GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search, using sw model

Run on: October 2, 2003, 17:49:28 ; Search time 70 Seconds
(without alignments)
1412.616 Million cell updates/sec

Title:          US-10-029-347-2
Perfect score:  3218
Sequence:       1 MLAQPQRLLFILDGADELPA.........ITHPALDGHPQPKELISTF 625

Scoring table:  BLOSUM62
                Gapop 10.0 , Gapext 0.5

Searched:       587654 seqs, 158212981 residues

Total number of hits satisfying chosen parameters:    587654

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Published Applications AA:*
1: /cgn2_6/ptodata/2/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/2/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/2/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/2/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/2/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/2/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/2/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/2/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/2/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/2/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/2/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/2/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/2/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/2/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/2/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/2/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/2/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/2/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Match Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 3098 | 96.3 | 947 | 11 | US-09-965-621-18 | Sequence 18, Appl |
| 2 | 1247 | 38.8 | 317 | 11 | US-09-965-621-60 | Sequence 60, Appl |
| 3 | 1084 | 33.7 | 290 | 11 | US-09-965-621-61 | Sequence 61, Appl |
| 4 | 825.5 | 25.7 | 1061 | 15 | US-10-066-521-18 | Sequence 18, Appl |
| 5 | 825 | 25.6 | 1035 | 11 | US-09-965-621-24 | Sequence 24, Appl |
| 6 | 800.5 | 24.9 | 1034 | 12 | US-10-028-374-18 | Sequence 18, Appl |
| 7 | 800.5 | 24.9 | 1034 | 12 | US-10-132-967-5 | Sequence 5, Appl |
| 8 | 800.5 | 24.9 | 1034 | 12 | US-10-183-770-18 | Sequence 18, Appl |
| 9 | 800.5 | 24.9 | 1034 | 14 | US-10-127-516-5 | Sequence 5, Appl |
| 10 | 800.5 | 24.9 | 1034 | 14 | US-10-027-629-5 | Sequence 5, Appl |
| 11 | 762 | 23.7 | 1429 | 10 | US-09-996-617-2 | Sequence 2, Appl |
| 12 | 762 | 23.7 | 1429 | 10 | US-09-931-071-2 | Sequence 2, Appl |
| 13 | 762 | 23.7 | 1429 | 12 | US-10-028-374-15 | Sequence 15, Appl |
| 14 | 762 | 23.7 | 1429 | 12 | US-10-183-770-15 | Sequence 15, Appl |
| 15 | 762 | 23.7 | 1429 | 15 | US-10-028-392-11 | Sequence 11, Appl |
| 16 | 762 | 23.7 | 1454 | 10 | US-09-388-221-10 | Sequence 10, Appl |
| 17 | 762 | 23.7 | 1473 | 10 | US-09-388-221-2 | Sequence 2, Appli |
| 18 | 760 | 23.6 | 1429 | 12 | US-10-028-374-3 | Sequence 3, Appl |
| 19 | 760 | 23.6 | 1429 | 12 | US-10-183-770-3 | Sequence 3, Appli |
| 20 | 716 | 22.2 | 1399 | 10 | US-09-388-221-4 | Sequence 4, Appl |
| 21 | 716 | 22.2 | 1424 | 10 | US-09-388-221-12 | Sequence 12, Appl |
| 22 | 716 | 22.2 | 1443 | 10 | US-09-388-221-6 | Sequence 6, Appl |
| 23 | 685 | 21.3 | 719 | 11 | US-09-965-621-59 | Sequence 59, Appl |
| 24 | 650 | 20.2 | 980 | 9 | US-09-848-035-13 | Sequence 13, Appl |
| 25 | 650 | 20.2 | 980 | 12 | US-09-986-224-13 | Sequence 13, Appl |
| 26 | 641 | 19.9 | 449 | 12 | US-10-028-374-2 | Sequence 2, Appl |
| 27 | 641 | 19.9 | 449 | 12 | US-10-183-770-2 | Sequence 2, Appl |
| 28 | 634 | 19.7 | 1049 | 12 | US-10-239-663-42 | Sequence 42, Appl |
| 29 | 619.5 | 19.3 | 1033 | 12 | US-10-132-967-2 | Sequence 2, Appl |
| 30 | 619.5 | 19.3 | 1033 | 14 | US-10-127-516-2 | Sequence 2, Appl |
| 31 | 619.5 | 19.3 | 1033 | 14 | US-10-027-629-2 | Sequence 2, Appl |
| 32 | 616.5 | 19.2 | 1062 | 12 | US-10-239-663-43 | Sequence 43, Appl |
| 33 | 598.5 | 18.6 | 858 | 9 | US-09-848-035-8 | Sequence 8, Appl |
| 34 | 598.5 | 18.6 | 858 | 10 | US-09-986-224-8 | Sequence 8, Appl |
| 35 | 598.5 | 18.6 | 994 | 11 | US-09-965-621-16 | Sequence 16, Appl |
| 36 | 598.5 | 18.6 | 994 | 15 | US-10-066-521-24 | Sequence 24, Appl |
| 37 | 597.5 | 18.6 | 1016 | 10 | US-09-996-224-19 | Sequence 19, Appl |
| 38 | 586.5 | 18.2 | 896 | 15 | US-10-066-521-22 | Sequence 22, Appl |
| 39 | 567.5 | 17.6 | 514 | 12 | US-10-028-374-14 | Sequence 14, Appl |
| 40 | 567.5 | 17.6 | 514 | 12 | US-10-183-770-14 | Sequence 14, Appl |
| 41 | 567.5 | 17.6 | 514 | 15 | US-10-028-392-10 | Sequence 10, Appl |
| 42 | 511.5 | 15.9 | 1143 | 15 | US-10-216-645-4 | Sequence 4, Appl |
| 43 | 511.5 | 15.9 | 1162 | 15 | US-10-216-645-2 | Sequence 2, Appl |
| 44 | 511.5 | 15.9 | 1344 | 15 | US-10-066-521-6 | Sequence 6, Appl |
| 45 | 501.5 | 15.6 | 821 | 9 | US-09-848-035-2 | Sequence 2, Appl |

ALIGNMENTS

RESULT 1
US-09-965-621-18

; Sequence 18, Application US/0996521
; Publication No.US20030077699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztcf
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
; LENGTH: 947
; TYPE: PRT
; ORGANISM: Homo sapien
US-09-965-621-18

Query Match             96.3%;
Best Local Similarity   90.0%;
Matches  613; Conservative    1; Mismatches   11; Indels   56; Gaps    2;

Qy    1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  267 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 326

Qy   61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFDERRAERAYRFVKENETLFALCF 120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  327 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFDERRAERAYRFVKENETLFALCF 386

Converted to 60/961,361

```
Qy 121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 387 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 446

Qy 181 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 240
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 447 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 506

Qy 241 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 300
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 507 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 566

Qy 301 VSERVKQEALRWVQGQGQGCFGVAPEVTEGAKGLEDTEEPEEEEEGEEPNYPLELLYCLY 360
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 567 VSERVKQEALRWVQGQGQGCFGVAPEVTEGAKGLEDTEEPEEEEEGEEPNYPLELLYCLY 626

Qy 361 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 627 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 686

Qy 421 KKSLGKRLQASLGGGSSQGT---------------------------------------- 445
       ||||||||||||||||||||
Db 687 KKSLGKRLQASLGGGSSWLGTQLAPEVPFRPPCCDICPTPPPDRLLQGKAFARVPLNIAP 746

Qy 441 TKQLPASLL------------------------------------HPLFQAMTDPLCHLSSLTLSHCKLPAVCRDLSEA 484
       |||                                           |||||||||||||||||||||||||||||
Db 747 IQPLPRGLASVERMNVTVLAGAGPGDPKTHAVTDPLCHLSSLTLSHCKLPDAVCRDLSEA 806

Qy 485 LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQS 544
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 807 LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQS 866

Qy 545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 867 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 926

Qy 605 VITHPALDGHPQPPKELISTF 625
       |||||||||||||||||||||
Db 927 VITHPALDGHPQPPKELISTF 947
```

RESULT 2
US-09-965-621-60
; Sequence 60, Application US/0996S621
; Publication No. US2003007699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 60
; LENGTH: 317
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-621-60

Query Match      38.8%;  Score 1247;  DB 11;  Length 317;
Best Local Similarity  100.0%;  Pred. No. 1.7e-110;
Matches 246;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```
Qy   : MLAQPQRLLFIILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 72  MLAQPQRLLFIILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 131

Qy 61  RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFALCF 120
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 132 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFALCF 191

Qy 121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 180
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 192 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 251

Qy 181 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 240
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 252 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 311

Qy 241 AALSYL 246
       ||||||
Db 312 AALSYL 317
```

RESULT 3
US-09-965-621-61
; Sequence 61, Application US/0996S621
; Publication No. US2003007699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 61
; LENGTH: 290
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-965-621-61

Query Match      33.7%;  Score 1084;  DB 11;  Length 290;
Best Local Similarity  76.6%;  Pred. No. 5.6e-95;
Matches 222;  Conservative 2;  Mismatches 1;  Indels 56;  Gaps 2;

```
Qy 392 VLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQASLGGGSSQGT------------ 440
       ||||||||||||||||||||||||||||||||||||||||||||||||
Db 1   VLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQASLGGGSWLGTQLAPEVPFRPP 60

Qy 441 -----------------------------------TKQLPASLL 455
                                          |||
Db 61  CCDICPTPPPDRLLQGKAFARVPLNIAPIQPLPRGLASVERMNVTVLAGAGPGDPKTHA 120

Qy 456 MTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAW 515
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 121 MTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAW 180

Qy 516 PQCRVQTVRVQLPDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCG 575
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 181 PQCRVQTVRVQLPDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCG 240

Qy 576 LQTLSLASVELSEQSLQELQAVKRAKPDLVITHPALDGHPQPPKELISTF 625
       |||||||||||||||||||||||||||||||||||||||||||||||||
Db 241 LQTLSLASVELSEQSLQELQAVKRAKPDLVITHPALDGHPQPPKELISTF 290
```

RESULT 4
US-10-066-521-18
; Sequence 18, Application US/10066521
; Publication No. US20030027757A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  APPLICANT: Wang, Weiye
;  APPLICANT: Blatcher, Maria
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN/NBS/LRR
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-334001
;  CURRENT APPLICATION NUMBER: US/10/066,521
;  CURRENT FILING DATE: 2002-06-25
;  PRIOR APPLICATION NUMBER: 60/318,645
;  PRIOR FILING DATE: 2001-09-10
;  PRIOR APPLICATION NUMBER: 60/265,231
;  PRIOR FILING DATE: 2001-01-31
;  NUMBER OF SEQ ID NOS: 25
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 18
;  LENGTH: 1061
;  TYPE: PRT
;  ORGANISM: Homo sapiens
US-10-066-521-18

Query Match             25.7%;  Score 825.5;  DB 15;  Length 1061;
Best Local Similarity   32.6%;  Pred. No. 2.1e-69;
Matches 219;  Conservative 104;  Mismatches 245;  Indels 103;  Gaps 15;

Qy      1 MLAQPQRLLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 59
            ::   :|:||||||:|| ||| |||   : :  :: ||::  |: |: || ||:|| ||:
Db    282 LIRVPERLLFIIDGFDELKPSFHDPQGPWCLCWEEKRPTELLLNSLIRKLLPELSLLIT 341

Qy     60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
           ||||| ||  |::|||:|:|:| : ||:|::: :::|:|||:::||::|:||:||||:||
Db    342 TRPTALEKLHRLLEHPRHVEILGFSEAERKEYFYKYFHNAEQAGQVFNYVRDNEPLFTMC 401

Qy    120 FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLQ--GDLR 177
           ||| ||||||::|||||| |||||| || ||||:|:|:|||:|::: :|: ||  |::
Db    402 FVPLVCWVVCTCLQQQLEGGGILRQTSRTTTAVIMLYLLSLMQPKPGA--PRLQPPNQR 459

Qy    178 NLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQ 237
           |:|| || ::||: ::::||| ||:|:::::::::|:||:|||||:::|::|:||:|||
Db    460 GLCSLAADGLWNQKILFEEQDLRKHGLDGEDV-SAFLMNIFQKDINCERYYSFIHLSFQ 518

Qy    238 EFLAALSYLLEDGGVPRTAAGGVG------TLLRGDAQPSHLVLTTRFLFGLLSAERM 290
           |:|:| : : |:::|: :            ||||||  | ||::: |:|:||::|:
Db    519 EFFAAMYYILDEG------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLFGLLNEETR 572

Qy    291 RDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPN 350
           ||:||||||||||||||||||||||||||||||||||||||||:||::||||||||||
Db    573 SHLEKSLCWKVSPHIKMDLLQWIQSKAQ----------------SDGSTLQQGS--- 610

Qy    351 YPLELLYCLYETQEDAFVRQALCRFPELA:QRVRFCRMDVAVLSYCVRCCPAGQALRLIS 410
           ||:|:|| ||||:|||:||||| ||||: : :  :: ::: || :: || || |||:|
Db    611 --LEFFSCLYEIQEEFIQQALSHFQVIVSNIA-SKMEHMVSSFCLKRCRSAQVLHLYG 667

Qy    411 CRLVAAQEKKKKKSLGKRLQASLGGGS---------------SQGTTKQLPASL--- 448
           || |:||:::::::  :|  :  |||:                 :::| |:
Db    668 ATYSADGE------DRARCSAGAHTLLVQLPERTVLLDAYSEHLAAALCTNPNLIELS 720

Qy    449 --------LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLH 499
                   |||:|||||:|||:|:|:||||:||:|| ||:|::||:| ||| ||:|
Db    721 YRNALGSRGVKLLCQGLRHPCRLQMIQLRKCQLESGACQEMASVLGTNPHLVELD:TGNALE 780

Qy    500 RLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTTLDLSGCQLP 558
           |:|:||||||||||||||||||||||||||||| ||||||||||||||||||||||||
Db    781 GVGFPGMLLCEGLRHPQCRLQMIQLRKCQLESGACQEMASVLGTNPHLVELD:TGNALE 840

Qy    559 APMVTYLCAVLQHQGCGLQTL-----------SLASVELSEQSLQELQAVKRAKPDLV 605
           || |||| |:||||| |||||            |:::|::||::|:|:|:|||:||:|
Db    841 DLGLRLLCQGLRHPVCRLRTLWLKIKICRLTAAACDELASTLSVNQSLRELDLSLNELGDLG 900

Qy    606 I-------THP 609
                   |||
Db    901 VLLLCEGLRHP 911

---

RESULT 5
US-09-965-621-24
; Sequence 24, Application US/09965621
; Publication No.US20030077699A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, John C.
;  APPLICANT: Godzik, Adam
;  APPLICANT: Chu, Zhi-Liang
;  APPLICANT: Pawlowski, Krzysztof
;  APPLICANT: Fiorentino, Loredana
;  APPLICANT: Ariza, Maria Eugenia
;  APPLICANT: Stehlik, Christian
;  TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
;  TITLE OF INVENTION:  Encoding Nucleic Acids, and Methods of Use
;  FILE REFERENCE: P-LJ 4816
;  CURRENT APPLICATION NUMBER: US/09/965,621
;  CURRENT FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 09/671,760
;  PRIOR FILING DATE: 2000-09-26
;  NUMBER OF SEQ ID NOS: 64
; SEQ ID NO 24
;  LENGTH: 1035
;  TYPE: PRT
;  ORGANISM: Homo sapien
US-09-965-621-24

Query Match             25.6%;  Score 825;  DB 11;  Length 1035;
Best Local Similarity   32.6%;  Pred. No. 2.3e-69;
Matches 219;  Conservative 104;  Mismatches 245;  Indels 104;  Gaps 15;

Qy      1 MLAQPQRLLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 59
            ::   :|:||||||:|| ||| |||   : :  :: ||::  |: |: || ||:|| ||:
Db    255 LIRVPERLLFIIDGFDELKPSFHDPQGPWCLCWEEKRPTELLLNSLIRKLLPELSLLIT 314

Qy     60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
           ||||| ||  |::|||:|:|:| : ||:|::: :::|:|||:::||::|:||:||||:||
Db    315 TRPTALEKLHRLLEHPRHVEILGFSEAERKEYFYKYFHNAEQAGQVFNYVRDNEPLFTMC 374

Qy    120 FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLQ--GDLR 177
           ||| ||||||::|||||| |||||| || ||||:|:|:|||:|::: :|: ||  |::
Db    375 FVPLVCWVVCTCLQQQLEGGGLLRQTSRTTTAVYMLYLLSLMQPKPGA--PRLQPPNQR 432

Qy    178 NLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQ 237
           |:|| || ::||: ::::||| ||:|:::::::::|:||:|||||:::|::|:||:|||
Db    433 GLCSLAADGLWNQKILFEEQDLRKHGLDGEDV-SAFLMNIFQKDINCERYYSFIH:SFQ 491

Qy    238 EFLAALSYLLEDGGVPRTAAGGVG------TLLRGDAQPSHLVLTTRFLFGLLSAERM 290
           |:|:| : : |:::|: :            ||||||  | ||::: |:|:||::|:
Db    492 EFFAAMYYILDEG------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLFGLLNEETR 545

Qy    291 RDERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPN 350
           ||:||||||||||||||||||||||||||||||||||||||||:||::||||||||||
Db    546 SHLEKSLCWKVSPHIKMDLLQWIQSKAQ----------------SDGSTLQQGS--- 583

Qy    351 YPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLIS 410
           ||:|:|| ||||:|||:||||| ||||: : :  :: ::: || :: || || |||:|
Db    584 --LEFFSCLYEIQEEFIQQALSHFQVIVSNIA-SKMEHMVSSFCLKRCRSAQVLHLYG 640

Qy    411 CRLVAAQEKKKKKSLGKRIQASLGGGS---------------SQGTTKQLPASL--- 448
           || |:||:::::::  :|  :  |||:                 :::| |:
Db    641 ATYSADGE------DRARCSAGAHTLLVQLRPERTVLLDAYSEHLAAALCTNPNLIELS 693

Qy    449 --------LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLH 498
                   |||:|||||:|||:|:|:||||:||:|| ||:|::||:| ||| ||:|
Db    694 LYRNALGSRGVKLLCQGLRHPCNCKLQNLRLKRCRISSSACEDLSAALIANKNLTRMDLSG 753

Qy    499 NRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTTLDLSGCQL 557

```
Db  754 NGVGFPGVMLLCEGLRHPQCRLQMIQLRKCQLESGACQEMASVLGTNPHLVELDLTGNAL 813
Qy  558 PAPMVTYLCAVLQHGCGLQTL------------SLASVELSEQSLQELQAVKRAKPDL 604
Db  814 EDLGLRLLCQGLRHPVCRLRTLWLKICRLTAAACDELASTLSVNQSLRELDLSLNELGDL 873
Qy  605 V:------THP 609
Db  874 GVLLCEGLRHP 885
```

RESULT 6
US-10-028-374-18
; Sequence 18, Application US/10028374
; Publication No. US2003014370641
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED
; TITLE OF INVENTION: PREDOMINATELY IN BONE MARROW, HLRRBM1
; FILE REFERENCE: D0067NP
; CURRENT APPLICATION NUMBER: US/10/028,374
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/257,773
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 22
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 18
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-028-374-18

Query Match        24.9%;   Score 800.5;  DB 12;  Length 1034;
Best Local Similarity   28.4%;   Pred. No. 5.1e-67;
Matches  219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps 17;

```
Qy    1 MLAQPQRLFLFLDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLLVT 59
Db  288 IVRKPSRILFLMDGFDELQGQAFDEHIGPLCTDWQKAERGDILLSSLIRKKLLPEASLLIT 347

Qy   60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
Db  348 TRPVALEKLQHLLDHPRHVEILGFSEAKREYFFKYSDEAQARAAFSLIQENEVLFTMC 407

Qy  120 FVPFVCWIVCTVLRQQLEGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGQLRNL 179
Db  408 FIPLVCWIVCTGLKQQMESGKSLAQTSKTTTAVYVFFLSSLLQPRGGSQEHGLCAHLWGL 467

Qy  180 CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
Db  468 CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRMNLFQXEVDCKFYSPIHMTFQEF 526

Qy  240 LAALSYLLEDGGVPRTAAGG------------VGTLLRGDAQ-PHSHLVLTRFLFGLLSAER 289
Db  527 FAAMYYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER 586

Qy  290 MRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEGEP 349
Db  587 TSYLEKKLSCKISQQIRLELLKWI------EVKAKAKKLQ--IQPSQ 625

Qy  350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
Db  626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG 680

Qy  410 SCRLVAAQEKKKKSLGKRLQ------ASLGGGSSQG-TTKQLPASLLHPLF------QA 455
Db  681 FLHNMPKEEEEEKEGRHLDKVQCVLPSSSHAACSHGLVNSLTSSFCRGLFSVLSTSQS 740

Qy  456 MT:------DP 459
Db  741 LTELDLSDNSLGDPGMRVLCFTLQHPGCNIRLKLWLGRCGLSHECCFDISLVLSSNQKIVE 800
```

```
Qy  460 ------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL 496
Db  801 LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV 860

Qy  497 LHNRLSEAGLRMLSEGLAWPQCRVQTVRV------- 525
Db  861 GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSSVLSTNQNLTHLYLRGN 920

Qy  526 QLPDPQRGLQYLV-GMLRQSPALTIDLSGCQLAPM---- 561
Db  921 TLGD--KGIKLLCEGLLHPDCKLQVTLEDNCNLTSHCCWDLSTLLTSSQSLRKLLSLGNND 978

Qy  562 ------VTYLCAVLQHGCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI 606
Db  979 LGDLGVMMFCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPELTV 1028
```

RESULT 7
US-10-132-967-5
; Sequence 5, Application US/10132967
; Publication No. US20030170841A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THE
; FILE REFERENCE: 07334-311001
; CURRENT APPLICATION NUMBER: US/10/132,967
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: US/09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-132-967-5

Query Match        24.9%;   Score 800.5;   Pred. No. 5.1e-67;   Length 1034;
Best Local Similarity   28.4%;
Matches  219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps 17;

```
Qy    1 MLAQPQRLLFLDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLLVT 59
Db  288 IVRKPSRILFLMDGFDELQGQAFDEHIGPLCTDWQKAERGDILLSSLIRKKLLPEASLLIT 347

Qy   60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
Db  348 TRPVALEKLQHLLDHPRHVEILGFSEAKREYFFKYSDEAQARAAFSLIQENEVLFTMC 407

Qy  120 FVPFVCWIVCTVLRQQLEGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGQLRNL 179
Db  408 FIPLVCWIVCTGLKQQMESGKSLAQTSKTTTAVYVFFLSSLLQPRGGSQEHGLCAHLWGL 467

Qy  180 CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
Db  468 CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRMNLFQXEVDCKFYSPIHMTFQEF 526

Qy  240 LAALSYLLEDGGVPRTAAGG------------VGTLLRGDAQ-PHSHLVLTRFLFGLLSAER 289
Db  527 FAAMYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER 586

Qy  290 MRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEGEP 349
Db  587 TSYLEKKLSCKISQQIRLELLKWI------EVKAKAKKLQ--IQPSQ 625

Qy  350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
Db  626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG 680

Qy  410 SCRLVAAQEKKKKSLGKRLQ------ASLGGGSSQG-TTKQLPASLLHPLF------QA 455
```

Db    681  FLHNMPKEEEEKEGRHLDMVQCVLPSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS  740

Qy    456  MT------------DP-----------------------------------------  459

Db    741  LTELDLSDNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE  800

Qy    460  ------------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL  496
Db    801  LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV  860

Qy    497  LHNRLSEAGLRMLSEGLAWPQCRVQTVRV------------  525
Db    861  GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSSVLSTNQNLTHLYLRGN  920

Qy    526  QLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM------------  561
Db    921  TLGD--KGIKLLCEGLLHPDCKLQVLELDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND  978

Qy    562  -----VTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI  606
Db    979  LGDLGVMMFCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPELTV  1028

RESULT 8
US-10-183-770-18
; Sequence 18, Application US/10183770
; Publication No. US20030180812A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED
; TITLE OF INVENTION: PREDOMINATELY IN BONE MARROW, HLRRBM1
; FILE REFERENCE: D0067A CIP
; CURRENT APPLICATION NUMBER: US/10183770
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: US 60/257,773
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: US 10/228,374
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 18
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-183-770-18

Query Match      24.9%;  Score 800.5;  DB 12;  Length 1034;
Best Local Similarity   28.4%;  Pred. No. 5.1e-67;
Matches 219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps  17;

Qy      1  MLAQPQRLFILDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLVT  59
Db    288  IVRKPSRILFLMDGFDELQGAFDEHIGPLCTDWQKAERGDILLSSLIRKLLPEASLLIT  347

Qy     60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFALC  119
Db    348  TRPVALEKLQHLLDHPRHVEILGFSEAKRKEYFKYFSDEAQARAAFSLIQENEVLFTMC  407

Qy    120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNL  179
Db    408  FIPLVCWIVCTGLKQQMESGKSLAQTSKTTTAVVVFFLSSLLQPRGSQEHGLCAHLWGL  467

Qy    180  CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSSKKELPGVLETEVTYQFIDQSFCEF  239
Db    468  CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRMNLFQKEVDCEKFYSFIHVTFQEF  526

Qy    240  LAALSYLLEDGGVPRTAAGG--------VGTLLRGDAQ-FHSHLVLTTRFLFGGLLSAER  289
Db    527  FAAMYYLLEEEKEGRTNVPGSRLKLPSRDVTVLENYGKFEKGYLIFVVRFLFGGLVNQER  586

Qy    290  MRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTGEAKGLEDTEEPEEESEGEEP  349
Db    587  TSYLEKKLSCKISQQIRLELLKWI---------EVKAKAKKLQ--IQPSQ--------  625

Qy    350  NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI  409
Db    626  ---LELFYCJYEMQEEDFVQRAMDYPFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG  680

Qy    410  SCRLVAAQEKKKKSLGKKLQ------ASLGGGSSQG-TTKQLPASLLHPLF------QA  455
Db    681  FLHNMPKEEEEEKEGRHLDMVQCVLPSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS  740

Qy    456  MT------------DP-----------------------------------------  459
Db    741  LTELDLSDNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE  800

Qy    460  ------------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL  496
Db    801  LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV  860

Qy    497  LHNRLSEAGLRMLSEGLAWPQCRVQTVRV------------  525
Db    861  GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSSVLSTNQNLTHLYLRGN  920

Qy    526  QLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM------------  561
Db    921  TLGD--KGIKLLCEGLLHPDCKLQVLELDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND  978

Qy    562  -----VTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI  606
Db    979  LGDLGVMMFCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPELTV  1028

RESULT 9
US-10-127-516-5
; Sequence 5, Application US/10127516
; Publication No. US20020187922A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Manji, Gulam A.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-370001
; CURRENT APPLICATION NUMBER: US/10/127,516
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 10/027,629
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/964,955
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506, 067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
; LENGTH: 1034
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-127-516-5

Query Match      24.9%;  Score 800.5;  DB 14;  Length 1034;
Best Local Similarity   28.4%;  Pred. No. 5.1e-67;
Matches 219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps  17;

Qy      1  MLAQPQRLFILDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLVT  59
Db    288  IVRKPSRILFLMDGFDELQGAFDEHIGPLCTDWQKAERGDILLSSLIRKLLPEASLLIT  347

Qy     60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFALC  119
Db    348  TRPVALEKLQHLLDHPRHVEILGFSEAKRKEYFKYFSDEAQARAAFSLIQENEVLFTMC  407

Qy    120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNL  179
Db    408  FIPLVCWIVCTGLKQQMESGKSLAQTSKTTTAVVVFFLSSLLQPRGSQEHGLCAHLWGL  467

```
Qy 180 CRLAREGVLGRRAQFAEKELFQELERGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
       |||.|:.|||..| .||:|||.|.|.||..|||| .||||||.||.| |.|||:||||
Db 468 CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRXNLFQKEVDCEKFYSFIHMTFQEF 526
Qy 240 LAALSYLLEDGGVPRTAAGG-------VGTLLRGDAQ-PHSHLVLTTRFLFGLLSAER 289
       :|| |:|||:.:: .: :|     : |||||||:|| :: ::..: ||||||::.||
Db 527 FAAMYYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER 586
Qy 290 MRDIERHFGCMVSERVKQEAIRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEP 349
       .::||:.:.| :|| ::|| |                ||.|:|| ::.|:.
Db 587 TSYLEKKLSCKISQQIRLELLKWI------EVKAXAKKLQ--IQPSQ-- 625
Qy 350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
       :.|||||||||.||::||.||| .||| | :|   | ||::.||::||:|:  : :.||
Db 626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG 680
Qy 410 SCRLVAAQEKKKSLGKRLQ--------ASLGGGSSQG-TTKQLPASLLHPLF-------QA 455
       . |:  :|:|: |||::||        ||.|.||:.| ||||:.:|:|:|| |       |:
Db 681 FLHNMPKEEEEEEKEGRHLDMVQCVLPSSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS 740
Qy 456 MT-------DP------ 459
       .|       |
Db 741 LTELDLDSNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE 800
Qy 460 -------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL 496
              |||::|||||||||||||||||||||||||||||||||
Db 801 LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV 860
Qy 497 LHNRLSEAGLRMLSEGLAWPQCRVQTVRV---- 525
       ||||||||||||||||||||||||||||
Db 861 GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSSVLSTNQNLTHLYLRGN 920
Qy 526 QLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM---------- 561
       |||||||||||| |||||||||||||||||||||||
Db 921 TLGD--KGIKLLCEGLLHPDCKLQVLELDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND 978
Qy 562 -----VTYLCAVLQHQCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI 606
            ||||||||||||||||||||||||||||||||||||||||||||
Db 979 LGDLGVMMFCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPELTV 1028
```

Query Match    24.9%;  Score 800.5;  DB 14;  Length 1034;
Best Local Similarity    28.4%;  Pred. No. 5.1e-67;
Matches 219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps 17;

```
Qy 1 MLAQPQRLLFILDGADEZPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLLVT 59
```

```
Db 288 IVRKPSRILFLMDGFDELQGAFDEHIGPLCTDWQKAERGDILLSSLIRKKLLPEASLLIT 347
       :: .| .:||.:||.||| |||.|:.||| ||| ::|.| |||||.||||.|||:|.|:|
Qy 60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
Db 348 TRPVALEKLQHLLDHPRHVEILGFSEAKRKEYFFKYFSDEAQARAAFSLIQENEVLFTMC 407
Qy 120 FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFITSVLLFFSSAPVADGPRLQGDLRNL 179
       |.|:||||||| ||||:|.|:.|.:||||||:|.:||||:||||.|||.|.:|| .|||:
Db 408 FIPLVCWIVCTGLKQQMESGKSLAQTSKTTTAVYVFFLSSLLQPRGGSQEHGLCALLWGL 467
Qy 180 CRLAREGVLGRRAQFAEKELFQELERGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
       |||.|:.|||..| .||:|||.|.|.||..|||| .||||||.||.| |.|||:||||
Db 468 CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRXNLFQKEVDCEKFYSFIHMTFQEF 526
Qy 240 LAALSYLLEDGGVPRTAAGG-------VGTLLRGDAQ-PHSHLVLTTRFLFGLLSAER 289
Db 527 FAAMYYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER 586
Qy 290 MRDIERHFGCMVSERVKQEAIRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEP 349
Db 587 TSYLEKKLSCKISQQIRLELLKWI------EVKAXAKKLQ--IQPSQ-- 625
Qy 350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
Db 626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG 680
Qy 410 SCRLVAAQEKKKSLGKRLQ--------ASLGGGSSQG-TTKQLPASLLHPLF-------QA 455
Db 681 FLHNMPKEEEEEEKEGRHLDMVQCVLPSSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS 740
Qy 456 MT-------DP------ 459
Db 741 LTELDLDSNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE 800
Qy 460 -------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL 496
Db 801 LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV 860
Qy 497 LHNRLSEAGLRMLSEGLAWPQCRVQTVRV---- 525
Db 861 GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSSVLSTNQNLTHLYLRGN 920
Qy 526 QLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM---------- 561
Db 921 TLGD--KGIKLLCEGLLHPDCKLQVLELDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND 978
Qy 562 -----VTYLCAVLQHQCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI 606
Db 979 LGDLGVMMFCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPELTV 1028
```

; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-996-617-2

Query Match       23.7%;  Score 762;  DB 10;  Length 1429;
Best Local Similarity  33.2%;  Pred. No. 4e-63;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

```
Qy     1 MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 58
           :||:||:||||||| :|  :|||:: |||||||||::|| |::|:||:|||||:|||:|
Db   398 ILSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPQPADALLGSLLGKTILPEASFLI 456

Qy    59 TTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFAL 118
           ||||||||:||:||||||:|||:|||::|:||||||||||||||:|||||:||||:||
Db   457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL 516

Qy   119 CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRN 178
           |||||||:||||:|||||:||||::||:|||||:|||||:||:|||||||| ||| ||:
Db   517 CLVPWVSWLACTCLMQQMRKEKLTLTSKTTTTLCLHYLAQALQAQPL--GPQ----LRD 570

Qy   179 LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETE---VTYQFIDQS 235
           :|||||||:|||:|| | ||:||||::|:| |||||||||| :  |:   | ||||:|
Db   571 LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM---GILQEHPIPLSYSFIHLC 625

Qy   236 FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL----------------T 277
           |||| ||||||||                    :KGRG------KHSNCIIDLEKTLEAYG.HGLFGAST
Db   626 FQEFFAAMSYVLED----------------EKGRG------KHSNCIIDLEKTLEAYG.HGLFGAST 670

Qy   278 TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVCQGQGCPGVAPEVTEGAKGLEDT 337
           ||||:||||||||:||: ||||                    :|: :||
Db   671 TRFLLGLLSDEGEREMENIFHCRLSQ-----GRNLMQWV---------------PSLQLL 710

Qy   338 EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC---RMDVAVLS 394
           :|    |||||||:|::| | :                    ||::
Db   711 LQP------------HSLESLHCLYETRNKTFLTQVMAHFEEMGX--------CVETDMELLVCT 755

Qy   395 YCVRCCPAGQALRLISCR---------------------------LVAAQEKKKKSLGKRLQ 429
           |||:||||||||||||:
Db   756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVLFRWVPVTDAYWQILFSVLKVTRNL-KELD 814

Qy   430 ASLGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLLTLSHCKLPDAVCRDLSEA.RAAP 489
           |||||:|||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   815 LS----GNS-----LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 865

Qy   490 ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTLRRVQLPDPQRGL---QYLVGMLRQS 544
           ||||||||||||||:|||||| ||| ||:||||:|:|||:||
Db   866 TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS 921

Qy   545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||:||
Db   922 PSLKELDLQQNNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy   605 VI 606
          
Db   982 LI 983
```

RESULT 12
US-09-931-071-2
; Sequence 2, Application US/09931071
; Patent No. US2002012219A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-335001
; CURRENT APPLICATION NUMBER: US/05/931,071
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/428,252
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 2
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-931-071-2

Query Match       23.7%;  Score 762;  DB 10;  Length 1429;
Best Local Similarity  33.2%;  Pred. No. 4e-63;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

```
Qy     1 MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 22
Db   398 ILSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPQPADALLGSLLGKTILPEASFLI 58

Qy    59 TTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAERAYRFVKENETLFAL 456
Db   457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL 118

Qy   119 CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRN 516
Db   517 CLVPWVSWLACTCLMQQMRKEKLTLTSKTTTTLCLHYLAQALQAQPL--GPQ----LRD 178

Qy   179 LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETE---VTYQFIDQS 570
Db   571 LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM---GILQEHPIPLSYSFIHLC 235

Qy   236 FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL------EKGRG------KHSNCIIDLEKTLEAYG.HGLFGAST 625
Db   626 FQEFFAAMSYVLED----------------EKGRG------KHSNCIIDLEKTLEAYG.HGLFGAST 277

Qy   278 TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVCQGQGCPGVAPEVTEGAKGLEDT 670
Db   671 TRFLLGLLSDEGEREMENIFHCRLSQ-----GRNLMQWV---------------PSLQLL 337

Qy   338 EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC---RMDVAVLS 710
Db   711 LQP------------HSLESLHCLYETRNKTFLTQVMAHFEEMGX--------CVETDMELLVCT 394

Qy   395 YCVRCCPAGQALRLISCR---------------------------LVAAQEKKKKSLGKRLQ 755
Db   756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVLFRWVPVTDAYWQILFSVLKVTRNL-KELD 429

Qy   430 ASLGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLLTLSHCKLPDAVCRDLSEA.RAAP 814
Db   815 LS----GNS-----LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 489

Qy   490 ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTLRRVQLPDPQRGL---QYLVGMLRQS 865
Db   866 TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS 544

Qy   545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 921
Db   922 PSLKELDLQQNNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 604

Qy   605 VI 606
Db   982 LI 983
```

RESULT 13
US-10-028-374-15
; Sequence 15, Application US/10028374
; Publication No. US20030143706A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED
; TITLE OF INVENTION: PREDOMINATELY IN BONE MARROW, HLRRBM1
; FILE REFERENCE: D0067NP
; CURRENT APPLICATION NUMBER: US/10/028,374
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 60/257,773
; PRIOR FILING DATE: 2000-12-22
; NUMBER OF SEQ ID NOS: 22

; SOFTWARE: Patentin version 3.0
; SEQ ID NO 15
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-028-374-15

Query Match             23.7%;  Score 762;  DB 12;  Length: 1429;
Best Local Similarity   33.2%;  Pred. No. 4e-63;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

```
Qy      1 MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 58
            ||:|:|||||||||||||  :|||||||||||||| :||||||||||:||||||||||:
Db    398 ILSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPCPADALLGSLLGKTILPEASFLI 456

Qy     59 TTRAAAPGRLGRLCSPQCAEVRGFSDKDKKKYFYKFFRDRRAERAYRFVKENETLFAL 118
            :|||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL 516

Qy    119 CFVPFVCWIVCTVLRQCLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRN 178
            |:|||||||||||||||:|||||||||||||||:|||||||||||||||||||||:||
Db    517 CLVPWVSWLACTCLMQQMKRKEKLTLTSKTTTLCLHYLAQALQAQPL--GPQ----LRD 570

Qy    179 LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETE--VTYQFIDQS 235
            |||||||||:|||||||| ||||||||||||||:||:||:|||||    |||||:||
Db    571 LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM----GILQEHPIPLSYSFIHLC 625

Qy    236 FQEFLAALSYILLEDGGVPRTAAGGVGTLLRGDAQPHSHLV- -----------------T 277
            |||||||||| :|||||||||||||||||||:||||||||
Db    626 FQEFFAAMSYILED--------------EKGRG-------KHSNCIIDLEKTLEAYGIHGLFGAST 670

Qy    278 TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDT 337
            ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||:||
Db    671 TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV--------------PSIQLL 710

Qy    338 EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC--RMDVAVLS 394
            |||  ||||||||:|||||||||||||:|:|||||||||||||    |:||:|||
Db    711 LQP---------HSLESLHCLYETRNKTFLTQVMAHFEEMGM-----CVETDMELLVCT 755

Qy    395 YCVRCCPAGQALRLISCR---------------------LVAAQEKKKKSLGKRLQ 429
            :|:| |:|:||||:||                         :||||||||:|||||:
Db    756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVLFRWVPVTDAYWQILFSVLKVTRNL-KELD 814

Qy    430 ASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP 489
            :|||| |  | :|:||||||||||||:|||||||||||:|||||||||||||||||||||
Db    815 LS------GNS------LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 865

Qy    490 ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGL------QYLVGMLRQS 544
            ||||||||||:||||||||||||||||||||||||||||||||      ||:|||||||
Db    866 TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS 921

Qy    545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
            ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||:|
Db    922 PSLKELDLQQNNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy    605 VI 606
            :|
Db    982 LI 983
```

; PRIOR APPLICATION NUMBER: US 10/028,374
; PRIOR FILING DATE: 2001-12-20
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: Patentin version 3.1
SEQ ID NO 15
; LENGTH: 1429
; TYPE: PRT
; ORGANISM: homo sapiens
US-10-183-770-15

Query Match             23.7%;  Score 762;  DB 12;  Length: 1429;
Best Local Similarity   33.2%;  Pred. No. 4e-63;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

```
Qy      1 MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 58
            ||:|:|||||||||||||  :|||||||||||||| :||||||||||:||||||||||:
Db    398 ILSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPCPADALLGSLLGKTILPEASFLI 456

Qy     59 TTRAAAPGRLGRLCSPQCAEVRGFSDKDKKKYFYKFFRDRRAERAYRFVKENETLFAL 118
            :|||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db    457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL 516

Qy    119 CFVPFVCWIVCTVLRQCLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRN 178
            |:|||||||||||||||:|||||||||||||||:|||||||||||||||||||||:||
Db    517 CLVPWVSWLACTCLMQQMKRKEKLTLTSKTTTLCLHYLAQALQAQPL--GPQ----LRD 570

Qy    179 LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETE--VTYQFIDQS 235
            |||||||||:|||||||| ||||||||||||||:||:||:|||||    |||||:||
Db    571 LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM----GILQEHPIPLSYSFIHLC 625

Qy    236 FQEFLAALSYILLEDGGVPRTAAGGVGTLLRGDAQPHSHLV- -----------------T 277
            |||||||||| :|||||||||||||||||||:||||||||
Db    626 FQEFFAAMSYILED--------------EKGRG-------KHSNCIIDLEKTLEAYGIHGLFGAST 670

Qy    278 TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDT 337
            ||||||||||||||||||||||||||||||||||||||:||||||||||||||||||:||
Db    671 TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV--------------PSIQLL 710

Qy    338 EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC--RMDVAVLS 394
            |||  ||||||||:|||||||||||||:|:|||||||||||||    |:||:|||
Db    711 LQP---------HSLESLHCLYETRNKTFLTQVMAHFEEMGM-----CVETDMELLVCT 755

Qy    395 YCVRCCPAGQALRLISCR---------------------LVAAQEKKKKSLGKRLQ 429
            :|:| |:|:||||:||                         :||||||||:|||||:
Db    756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVLFRWVPVTDAYWQILFSVLKVTRNL-KELD 814

Qy    430 ASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP 489
            :|||| |  | :|:||||||||||||:|||||||||||:|||||||||||||||||||||
Db    815 LS------GNS------LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 865

Qy    490 ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGL------QYLVGMLRQS 544
            ||||||||||:||||||||||||||||||||||||||||||||      ||:|||||||
Db    866 TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS 921

Qy    545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
            ||||||||||||||||||||||||:||||||||||||||||||||||||||||||||:|
Db    922 PSLKELDLQQNNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy    605 VI 606
            :|
Db    982 LI 983
```

RESULT 14
US-10-183-770-15
; Sequence 15, Application US/10183770
; Publication No. US2003C180812A1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED
; TITLE OF INVENTION: PREDOMINATELY IN BONE MARROW, HLRRBM1
; FILE REFERENCE: D0067A CIP
; CURRENT APPLICATION NUMBER: US/10/183,770
; CURRENT FILING DATE: 2002-06-27
; PRIOR APPLICATION NUMBER: US 60/257,773
; PRIOR FILING DATE: 2000-12-22

RESULT 15
US-10-028-392-11
; Sequence 11, Application US/10028392
; Publication No. US2003008734CA1
; GENERAL INFORMATION:
; APPLICANT: Bristol-Myers Squibb Company
; TITLE OF INVENTION: A NOVEL HUMAN LEUCINE-RICH REPEAT CONTAINING PROTEIN EXPRESSED
; TITLE OF INVENTION: PREDOMINATELY IN NERVOUS SYSTEM TISSUES, HLRRNS1
; FILE REFERENCE: D0085.np
; CURRENT APPLICATION NUMBER: US/10/028,392

```
;  CURRENT FILING DATE: 2001-12-20
;  PRIOR APPLICATION NUMBER: US 60/259,479
;  PRIOR FILING DATE: 2001-01-03
;  PRIOR APPLICATION NUMBER: US 60/260,616
;  PRIOR FILING DATE: 2001-01-09
;  NUMBER OF SEQ ID NOS: 63
;  SOFTWARE: Patentin version 3.0
SEQ ID NO 11
;  LENGTH: 1429
;  TYPE: PRT
;  ORGANISM: homo sapiens
US-10-028-392-11

Query Match        23.7%;  Score 762;  DB 15;  Length 1429;
Best Local Similarity  33.2%;  Pred. No. 4e-63;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps  22;

Qy    1 MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV  58
          :|:||||||:|:|||:||:|   :|:|:|::|||||:||::||||::|:||::::||||||
Db  398 :LSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPQPADALLGSLLGKTILPEASFLI 456

Qy   59 TTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENET_FAL 118
          ||::|||||:||:|:|:|::|:|||::|:||||||||:|:|||||||||||:|:||:|||
Db  457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL 516

Qy  119 CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLQGDLRN 178
          |:||||:|||::||:|:||::::||:||:|:|::|:||:|||||::|||||:||||||::
Db  517 CLVPWVSWLACTCLMQQMKRKEXLTLTSKTTTLCHYLAQALQAQPL--GPQ----LRD 570

Qy  179 LCRLAREGVLGRRAQFAEKELEOLELRGSKVQTLFLSKKELPGVLETE--VTYQFIDQS 235
          |:|:|||||::|||::|::|:|:|||:::||:||:|||::|||:|:|:   |:|||:|:|
Db  571 LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM----GILQEHPIPLSYSFIHLC 625

Qy  236 FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL--------------T 277
          |:|||:|:||::|:|                :||||:::|||            :
Db  626 FQEFFAAMSYVLED------EKGRG------KESNCIIDLEKTLEAYGIHGLFGAST 670

Qy  278 TRFLFGLLSAERMRDIERHFG-MVSERVKQEALRWVQGQGQGQCPGVAPEVTEGAKGLFDT 337
          |||::|||:|:||::||:||  :|::|:|:|:||              |:|   PSLQLL 710
Db  671 TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV---

Qy  338 EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC--RMDVAVLS 394
          :|:        ::|:::||:|||:                |:|:       |:|:
Db  711 LQP--------HSLES:HC:YETRNKTFLTQVMAHFEEMGM----CVETDMELLVCT 755

Qy  395 YCVRCCPAGQALRLISCR---------------------LVAAQEKKKKSJGKRLQ 429
          |::|:|:|                               ::::||||:|::|:|::
Db  756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVVLFRWVFVTDAYWQILFSVLKVTRNL-KELD 814

Qy  430 ASLGGGSQGTTKQLPASLLH:LFQAM:DPLCHLSSLT:SHCKLPDAVCRDLSEALRAAP 489
          |:             :|::|::|::|::||||::|:|::||||||||:|:||:|:|||:|
Db  815 LS-----GNS------LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 865

Qy  490 ALTELGLLHNRLSEAGLRMJSEGLAWPQCRVQTVRVQLPDDPQRGL----QYLVGMLRQS 544
          |::|||::|:|:||:|::|::||||::|:|:|||:|||:|:||        |:||::|:|
Db  866 T:TELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--R:QL--VSCGLTSDCQDLASVLSAS 92:

Qy  545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLCTLSLASVELSEQSLQEIQAVKRAKPDL 604
          |:::||||:||:|:|::|||:|::|:|||||:::||::||:||:::||:||:|::|:|::|
Db  922 PSLKELDLQQNNLDDVGVRLLCEGLRHPACKJIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy  605 VI 606
         :|
Db  982 LI 983
```

Search completed: October  2, 2003, 18:00:42
Job time : 75 secs

```
                    GenCore version 5.1.6
              Copyright (c) 1993 - 2003  Compugen Ltd.

CM protein - nucleic search, using frame_plus_p2n model

Run on:       October  3, 2003, 15:49:16 ; Search time 410 Seconds
                                            (without alignments)
                                            3888.511 Million cell updates/sec

Title:         US-10-029-347-2
Perfect score: 3218
Sequence:      1 MLAQPQRLFLDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table: BLOSUM62
               Xgapop   10.0 , Xgapext   0.5
               Ygapop   10.0 , Ygapext   0.5
               Fgapop    6.0 , Fgapext   7.0
               Delop     6.0 , Delext    7.0

Searched:      1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters:      3416838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
 -MODEL=frame+_p2n.model -DEV=xln
 -Q=/cgn2_1/USPTO_spool/US10029347/rrnat_02102003_173438_17:46/app_query.fasta_.775
 -DB=Published_Applications_NA -CFMT=fastap -SUFFIX=rnpb -MINMATCH=0.1
 -LOOPCL=0 -LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62
 -TRANS=human -LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100
 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0
 -MAXLEN=2000000000 -USER=US10029347 @CGN_1_221 @runat_02102003_173438_17146
 -NCPU=6 -ICPU=3 -NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100
 -LONGLOG -DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5
 -FGAPOP=6 -FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     Published_Applications_NA:*
     1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
     2: /cgn2_6/ptodata/1/pubpna/PCT_NEW_PUB.seq:*
     3: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
     4: /cgn2_6/ptodata/1/pubpna/US05_PUBCOMB.seq:*
     5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
     6: /cgn2_6/ptodata/1/pubpna/PCT5_PUBCOMB.seq:*
     7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
     8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
     9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
    10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
    11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
    12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
    13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
    14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
    15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
    16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
    17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                    SUMMARIES
Result   %
No.   Score  Match  Query Length  DB   ID                  Description
----  -----  -----  ------------  --   --                  -----------
  1   3098   96.3   2844     11    US-09-965-621-17        Sequence 17, Appl
```

| | | | | | | |
|---|---|---|---|---|---|---|
| 2 | 826.5 | 25.7 | 3108 | 11 | US-09-965-621-23 | Sequence 23, Appl |
| 3 | 825.5 | 25.7 | 3186 | 14 | US-10-066-521-17 | Sequence 17, Appl |
| 4 | 800.5 | 24.9 | 3102 | 12 | US-10-132-967-6 | Sequence 6, Appl |
| 5 | 800.5 | 24.9 | 3102 | 13 | US-10-127-516-6 | Sequence 6, Appl |
| 6 | 800.5 | 24.9 | 3102 | 13 | US-10-027-629-6 | Sequence 6, Appl |
| 7 | 800.5 | 24.9 | 3857 | 12 | US-10-132-967-4 | Sequence 4, Appl |
| 8 | 800.5 | 24.9 | 3857 | 13 | US-10-127-516-4 | Sequence 4, Appl |
| 9 | 800.5 | 24.9 | 3857 | 13 | US-10-027-629-4 | Sequence 4, Appl |
| 10 | 762 | 23.7 | 4287 | 10 | US-09-996-617-5 | Sequence 5, Appl |
| 11 | 762 | 23.7 | 4287 | 10 | US-09-931-071-5 | Sequence 5, Appl |
| 12 | 762 | 23.7 | 4422 | 10 | US-09-388-221-1 | Sequence 1, Appl |
| 13 | 762 | 23.7 | 4556 | 10 | US-09-388-221-9 | Sequence 9, Appl |
| 14 | 762 | 23.7 | 5444 | 10 | US-09-996-617-1 | Sequence 1, Appl |
| 15 | 762 | 23.7 | 5444 | 10 | US-09-931-071-1 | Sequence 1, Appl |
| 16 | 762 | 23.7 | 5444 | 11 | US-09-956-712-3 | Sequence 3, Appl |
| 17 | 762 | 23.7 | 6531 | 11 | US-09-956-712-11 | Sequence 12, Appl |
| 18 | 760 | 23.6 | 5100 | 11 | US-09-956-712-12 | Sequence 12, Appl |
| 19 | 716 | 22.2 | 4200 | 10 | US-09-388-221-3 | Sequence 3, Appl |
| 20 | 716 | 22.2 | 4332 | 10 | US-09-388-221-5 | Sequence 5, Appl |
| 21 | 716 | 22.2 | 4466 | 10 | US-09-388-221-11 | Sequence 11, Appl |
| 22 | 694 | 21.6 | 4931 | 12 | US-10-028-374-1 | Sequence 1, Appl |
| 23 | 694 | 21.6 | 4931 | 12 | US-10-183-770-1 | Sequence 1, Appl |
| 24 | 691.5 | 21.5 | 2524 | 11 | US-09-965-62-158 | Sequence 58, Appl |
| 25 | 650 | 20.2 | 2940 | 9 | US-09-848-035-14 | Sequence 14, Appl |
| 26 | 650 | 20.2 | 2940 | 10 | US-09-986-224-14 | Sequence 14, Appl |
| 27 | 650 | 20.2 | 3263 | 9 | US-09-848-035-12 | Sequence 12, Appl |
| 28 | 650 | 20.2 | 3263 | 10 | US-09-986-224-12 | Sequence 12, Appl |
| 29 | 634 | 19.7 | 3150 | 12 | US-10-239-663-9 | Sequence 9, Appl |
| 30 | 619.5 | 19.3 | 3099 | 12 | US-10-132-967-3 | Sequence 3, Appl |
| 31 | 619.5 | 19.3 | 3099 | 13 | US-10-127-516-3 | Sequence 3, Appl |
| 32 | 619.5 | 19.3 | 3099 | 13 | US-10-027-629-3 | Sequence 3, Appl |
| 33 | 619.5 | 19.3 | 3431 | 12 | US-10-132-967-1 | Sequence 1, Appl |
| 34 | 619.5 | 19.3 | 3431 | 13 | US-10-127-5-16-1 | Sequence 1, Appl |
| 35 | 619.5 | 19.3 | 3431 | 13 | US-10-027-629-1 | Sequence 1, Appl |
| 36 | 616.5 | 19.2 | 3189 | 12 | US-10-239-663-10 | Sequence 10, Appl |
| 37 | 598.5 | 18.6 | 2575 | 9 | US-09-848-035-7 | Sequence 7, Appl |
| 38 | 598.5 | 18.6 | 2575 | 10 | US-09-986-224-7 | Sequence 7, Appl |
| 39 | 598.5 | 18.6 | 2982 | 14 | US-10-066-521-25 | Sequence 25, Appl |
| 40 | 598.5 | 18.6 | 2985 | 11 | US-09-965-621-15 | Sequence 15, Appl |
| 41 | 598.5 | 18.6 | 3368 | 14 | US-10-066-521-23 | Sequence 23, Appl |
| 42 | 597.5 | 18.6 | 3048 | 10 | US-09-986-224-20 | Sequence 20, Appl |
| 43 | 597.5 | 18.6 | 3051 | 10 | US-09-986-224-18 | Sequence 18, Appl |
| 44 | 586.5 | 18.2 | 2691 | 14 | US-10-066-521-21 | Sequence 21, Appl |
| 45 | 566 | 17.6 | 96649 | 11 | US-09-956-712-10 | Sequence 10, Appl |

```
                                          ALIGNMENTS

RESULT 1
US-09-965-62-17
; Sequence 17, Application US/0965621
; Publication No. US2003007769A1
; GENERAL INFORMATION:
;  APPLICANT: Reed, John C.
;  APPLICANT: Godzik, Adam
;  APPLICANT: Chu, Zhi-Liang
;  APPLICANT: Pawlowski, Krzysztof
;  APPLICANT: Fiorentino, Loredana
;  APPLICANT: Ariza, Maria Eugenia
;  APPLICANT: Stehlik, Christian
;  TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
;  TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
;  FILE REFERENCE: P-LJ 4816
;  CURRENT APPLICATION NUMBER: US/09/965,621
;  CURRENT FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 09/671,760
;  PRIOR FILING DATE: 2000-09-26
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
;  LENGTH: 2844
;  TYPE: DNA
;  ORGANISM: Homo sapien
```

```
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2841)
US-09-965-621-17

Alignment Scores:
Pred. No.:           0
Score:            3098.00
Percent Similarity:  90.16%
Best Local Similarity: 90.01%
Query Match:         96.27%
DB:                  11

US-10-029-347-2 (1-625) x US-09-965-621-17 (1-2844)
```

```
Qy      1  MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799  ATGCTGGCCCAGCCGCAGCGGCTGTTCATCCTGGACGGGGCGGACGAGCTGCCGGCG  858

Qy     21  LeuGlyGlyProGluAlaAlaProCysThrArgAspPheGluAlaAlaSerGlyAlaArg  40
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    859  CTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCCCGG  918

Qy     41  ValLeuGlyGlyLeuLeuSerLysSerLeuProThrAlaLeuLeuProThrAlaThrThr  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    919  GTGCTAGGCGGGCTGCTGAGCAAGAGCCTGCCCACGGCGCCCTGCTGCTGTGACCACG  978

Qy     61  ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGluVal  80
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    979  CGGGCGGCCGCCCCGGGACGCCTGCAGGGCCGCCTGTGTCCCCGCAGTGCGCCGAGGTG  1038

Qy     81  ArgGlyPheSerAspLysSerLysLysTyrTrpPheTyrLysPhePheArgAspGluArg  100
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1039  CGCGGCTTCTCCGACAAGAGCAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGGAG  1098

Qy    101  ArgAlaGluArgGlyAlaTyrArgAlaPheValLysGluAsnGluThrLeuAlaLeuCysPhe  120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1099  AGGGCCGAGCGCGGCGCCTACCGCTTCGTGAAGGAGAACGAGACCCTGTTCGGCGTGCTTC  1158

Qy    121  ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluLeuGlyArg  140
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1159  GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCTGGAGCTGGGTCGG  1218

Qy    141  AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal  160
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1219  GACCTGTCGCGCACGTCCAAGACCACCAGTGTATACCTGCTTTTCATCACCAGCGT  1278

Qy    161  LeuSerSerAlaProValAlaAlaAspGlyValLeuGlyValLeuArgGlnGlyLeuArgAsnLeuCys  180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1279  CTGAGCTCGGCTCCGGTAGCCGACGGGGTCCTGGGCGTGCTGAGGCAGGGCCTGCGCAATCTGTGC  1338

Qy    181  GlnLeuGluLeuArgGlyValLeuGlyValAlaArgArgAlaGlnPheAlaGlyGluLysGluGlu  200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1339  CAGCTGGAGCTGCGCGGCGTGCTCGGCGTGGCGCGGCGGGCGCAGTTTGCCGGAGAAGGAACTGGAG  1398

Qy    201  GlnLeuGluLeuArgGlySerValLysThrLeuPheLeuSerLysSerLysLysAsnGlyAla  220
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1399  CAACTGGAGCTTCGTGGCTCCAAAACCCTGTTTCTCAGCAAAAAGGAGCTCCCG  1458

Qy    221  GlyValLeuGluThrGluValThrThrTyrArgGlnPheIleGlnSerPheGlnSerPheGlnGlnLeu  240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1459  GGGGTGCTGGAGACAGAGGTCACCACCTACCGCCAGTTCATCCAGAGCTTCCAGGAGTTCCTC  1518

Qy    241  AlaAlaLeuSerTyrLeuGluGluSerGlyValGlyValAspGlyValLeuGlyValGlyVal  260
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1519  GCGGCACTGTCCTACCTGGAGGAGAGCGGGGTGGGCGTGGACGGCGTGCTGGGCGTGGCGGTT  1578

Qy    261  GlyThrLeuArgGlyAspAlaGlnProHisSerHisSerLeuValLeuThrThrArgPhe  280
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1579  GGGACACTCCTGCGTGGGGACGCCCAGCCCCACAGCCACAGCCTGGTGCTCACCACGCGCTTC  1638

Qy    281  LeuPheLeuGlyLeuSerArgLeuArgMetArgAspArgIleGluArgHisPheGlyCysMet  300
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1639  CTCTTCCGGCTGCTGAGCCGGCTGCGGATGCGCGACCGCATCGAGCGCCACTTCGGCTGCATG  1698
```

```
Qy    301  ValSerGluAlaValValGlnGlnGlnArgTrpValGlnGlyGlnGlyGlyGlnGlyGlyCys  320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1699  GTTTCAGAGGCTGTGAAGCAGCAGAGGTGGGTGCAGGGTCAGGGACAGGGCTGC  1758

Qy    321  ProGlyValAlaProGluValThrGluGlyValAlaLysGluAspThrGluProPro  340
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1759  CCCGGAGTGGCACCAGAGGTGACCGAGGGGGTGGCCAAAGAGGACACCGAAGAGCCA  1818

Qy    341  GluGluGluGluGluGlyGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyr  360
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1819  GAGGAGGAGGAGGAGGGCGAGGAGCCCAACACCACCTGGAGTTGCTGTACTGCCTGTAC  1878

Qy    361  GluThrGlnGluAspLysProPheValArgGlnAlaLeuCysArgPheProGluAlaLeu  380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1879  GAGACCCAGGAGGACGCGGTTTGTGCCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCGCTG  1938

Qy    381  GlnArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysSerValArgCysCys  400
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1939  CAGCGAGTGCGCTTCTGCCGCATGGACGTGGCCGTGCTGTCCTGAGCTACGTGCTGGTGCTGC  1998

Qy    401  ProAlaGlyGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnGluLysLysLys  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1999  CCTGCTGGACAGCACTGCGGCTGATCAGCTGCAGATTGGTGCTGCGGCAGAGAGAAGAAG  2058

Qy    421  LysLysSerLeuGlyArgLeuGlnLeuAlaSerLeuGlyLeuGlySerSerGlnGlyThr  440
           ||| ||| : |||||||||||||||||||||| |||||
Db   2059  AAGAAGAGCCTGGGAAGCGGCTCCAGGCAGCCGTTCAGGTGCGCGCAGCTGCTGGGGACC  2118

Qy    440  ------------------------------------------------------------  440
                                                                       
Db   2119  CAACTGGCTCCAGAAGTACCCTTCGACCACCCTGCTGCTGCGACATCTGCCCACCACCTCCA  2178

Qy    440  ------------------------------------------------------------  440
                                                                       
Db   2179  CCAGACCCTCGGCTCCTCCAGGGCAAGGCTTTGCCAGAGTTCCTTGAATATAGCCCCA  2238

Qy    441  ThrLysGlnLeuProAlaSerLeuLeu-------------------------------  449
           : :| ||||
Db   2239  ATTCAGCCCCTGCCCAGGGGGCTTGGCATCTGTGAGGATGAATGCACCGTGTGGCA  2298

Qy    450  -------------HisProLeuPheGlnAlaMetThrAspProLeuCysHisSerLeuSer  464
                        ||| ||||||||||||||||||||||||
Db   2299  GGGGCTGGCCTGGGACCCCAAAGACCAATGACTGACTGCCCACTGTGCCATCTGAGC  2358

Qy    465  SerLeuThrLeuSerHisCysLysLeuLeuProAlaAspAlaValCysArgAspLeuSerGluAla  484
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||: |||||
Db   2359  AGCCTCACGCTGTCCCACTGCAAACTCCTGCCCGCGGCGGATGCGCTGTGCCGAGACCTTTCTGAGGCC  2418

Qy    485  LeuArgAlaAlaProAlaAlaLeuThrGluLeuLeuHisAsnArgLeuSerArgLeuSerGluAla  504
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2419  CTGAGGGCAGCCCCCGCCACTGAACTGCTGCACAACAGGCTCAGTAGGCTCAGTGAGGCA  2478

Qy    505  GlyLeuArgMetLeuSerGluGlyGlyLeuAlaTrpProGlnCysProSerArgGlnThrValArg  524
           ||| || ||||||||||||||||||||||||||||||||||||||||||||||
Db   2479  GGACTGCGTATGCTGAGTGAGGGGCCTAGCCTGGCCCCAGTGCCCCAGTCCCAGTCCCAGG  2538

Qy    525  ValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValLeuGlyMetLeuArgGlnSer  544
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2539  GTACAGCTGCCTGACCCCCAGCGAGGGCCTCCAGTACCTGGTGGGTATGCTTCGGCAGAGC  2598

Qy    545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr  564
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2599  CCTGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGCCCGCCCCCATGGTGACCTAC  2658

Qy    565  LeuCysAlaValLeuGlnHisGlnLnGlyCysGlyLeuGlnThrLeuSerLeuAlaSerVal  584
           || ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2659  CTCTGTGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGACCCTCAGTCTGGCCTCTGTG  2718

Qy    585  GluLeuSerGlyGlnSerLeuGlnLeuGlnLnAlaAlaValLysProAlaLysHisAspLeu  604
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2719  GAGCTGAGCGGCCAGAGCCTGCAGCTGCAGCAGGCTGCTGTGAAGCCAGCAAAGCACGATCTG  2778
```

RESULT 2

US-09-965-621-23

; Sequence 23, Application US/09965621
; Publication No. US20030077699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 23
; LENGTH: 3108
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(3105)

US-09-965-621-23

Alignment Scores:

| | | | |
|---|---|---|---|
| Pred. No.: | 3.46e-84 | Length: | 3108 |
| Score: | 826.50 | Matches: | 231 |
| Percent Similarity: | 48.00% | Conservative: | 105 |
| Best Local Similarity: | 33.00% | Mismatches: | 247 |
| Query Match: | 25.68% | Indels: | 19 |
| DB: | 11 | Gaps: | 17 |

US-10-029-347-2 (1-625) x US-09-965-621-23 (1-3108)

```
Score:                825.50          Matches:         230
Percent Similarity:   48.07%          Conservative:    106
Best Local Similarity: 32.90%         Mismatches:      247
Query Match:          25.65%          Indels:          118
DB:                   14              Gaps:            17

US-10-029-347-2 (1-625) x US-10-066-521-17 (1-3186)

Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeu---Pro    19
            ::|:|       ||| |:|   :||  |: |    :|:   :|||   :
Db    844 CTCATCCGAGTTCCCAGCCGCCTCCTTTTCATCATCGACGGTGATGAGCTCAAGCCT    903

Qy     20 AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAla   39
            :::     |||       || ||  :|||
Db    904 AGGCATGATGCTGCTTTGCGAGGGCCTCGGCACCCCAGTCAGGCTGCAGATGATTCA   963

Qy     40 ArgValLeuGlyGlyLeuLeuSerLeuAlaLeuLeuProThrAlaLeuLeuValThr     59
            |:|       :|:   ||:  :|||||:  ||: ||:::
Db    964 CTGCTTCTTAACAGCTTAATTCCGAGAGAGCTGCTCTCCCCTGAGCTATCTTTGCCATCACC  1023

Qy     60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu   79
            |||||
Db   1024 ACACGGCCCACGGCTTTGGAGAAGCTCCACCGTCTGGAGCACCCCAGGCATGTGAG  1083

Qy     80 ValArgGlyPheSerAspLysLysAspLysLysTyrPheTyrLysPhePheArgAspGlu   99
            :::   :|||||:|  ||:||||::  ||||::     |||||:|   |
Db   1084 ATCCTGGGCTTCTGAGGCAGAAAAAGGAGGAATGAATTCTACAAGTATTTCCACAATGCA  1143

Qy    100 ArgArgAlaGluArgA:aTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys  119
            :::   :|||                  ||:||:
Db   1144 GAGCGAGGCGGGCAAGTCTTCAATTACGTGGAGGACACGAGCCTCTTCACCATGTGC  1203

Qy    120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnIleLeuGluLeuGly  139
            ||:|    ::   |||:   |||||    :|||     |||||:|:||
Db   1204 TTCGTCCCCTGGTGTGCTGGGTGGTGTCACCTGCCCCAGCAGCTGAGGGTGGG  1263

Qy    140 ArgAspSerArgThrSerLysThrThrThrSerValTyrLeuLeuPheIleThrSer    159
            ||: ||| |::       :||::       ||:||: |||
Db   1264 GGGCGTGTGAGCAGAGCTCCAGGATGCATCCCGCGGCC  1323

Qy    160 ValLeuSerSerAlaProValAlaAspGlyProArgLeuGln------GlyAspLeuArg  177
            :::   ||:|    ||:    |||       :::::::
Db   1324 CTGATGCAACCCAAGCCGGGGGCC-----CCCGGCCCCAGCCCCCAACCACCAGAGA  1377

Qy    178 AsnLeuCysArgLeuAlaArgGluGlyValLeuGlyValArgArgAlaGlnPheAlaGluLys  197
            ::|||   ::         |||::|    |||||   ||:|
Db   1378 GGGCTGTGTCCTTGGCCGCAGATGGGCTCTGGAATCAGAAAATCTATTTGGAGGACGACAG  1437

Qy    198 GluLeuGluGlnLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeuSerLysLys  217
            ::|||   |||   |:::|:    |||       ||:||
Db   1438 GACCTTCCGGAAGCCGGGACCACGACGGGGAAGACGTC---TCTGCCTTCCTCAACATGAAC  1494

Qy    218 GluLeuProGlyValLeuGluThrGlyValThrTyrGlnValThryrGlnLeuAspGlnSerPheGln  237
            ::|||:   |||||: ||:|   |||||::::
Db   1495 ATCTTCCGAGAGGACATCAACTGTGAGGAGTACTACAGCTTCATCCACTTGAGTTTCCAG  1554

Qy    238 GluProLeuAlaAlaLeuSerTyrLeuGluLeuGluAspGlyGlyValProGlyValProArgThrAlaAla  257
            ::|||:   :|||:|:     :|:  ||:|||||:  |||:
Db   1555 GAATTCTTTGCAGCTATGTACTATATCCTGGACGAGGGG---------------GAG  1596

Qy    258 GlyGlyValGly-----------------------ThrLeuArgGlyAspAlaGlnPro  270
            ||:|||:|:                            |||
Db   1597 GGCGGGCAGGCCGGGCAGCCCAGGACGTGACCAGGCGTGTTGACCGAGTACGCGGTTTCTGAA  1656

Qy    271 HisSerHisLeuValLeuThrArgPheGlyLeuLeuSerAlaGluArgMet  290
            |||:||:::||| |:|    |||   |
Db   1657 AGGAGCTTCCTGGCCACTCCACCAGCCGCTTCCTGTTTGGACTCCTGAACAGGAGGACCAGG  1716

Qy    291 ArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnAlaAlaLeu  310
            ||: ||:|::   |||     |||
Db   1717 AGCCACCTGGAAGAGAGTTCTTGCTGAGAAGTCTCGCCGCACCAAGATGACCCTGTTG  1776

Qy    311 ArgTrpValGlnGlyGlnGlnGlyGlyCysProGlyValAlaProGluValThrGluGly  330
```

```
Db   2097 GGGCAGCCGGGGGGTGAAGCTGCTGCTCTGTCAAGACTCAGGAGACTCAGACCCCAACTGCAAACTTCA  2156
            ||| :::   |||    |||
Qy    464 rSerLeuThrHisLysCysSerLeuProGluProAspAlaValCysArgAspLeuSerGluAl    484
            ::::|||          ::::   |||
Db   2157 GAACCTGAGGCTGAAGAGAGTGCCCGCATCCCAGCTCAGCCTGGCAGGACCCTCTTGCTGCAGC  2216

Qy    484 aLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuHisArgArgLeuSerGluAl   504
            :::|:|       || |:
Db   2217 TCTCATAGCCAATAAGAATTTGACAAGGATGGATCTCAGTGGCAACGGCGGTTGGATTCCC  2276

Qy    504 aGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrValAr   524
            :|||   |:|||   :|||
Db   2277 AGGCATGATGCTGCTTTGCGAGGGCCTCGGCACCCCCAGTCAGGCTGCAGATGATTCA  2336

Qy    524 gValGlnLeuProAspProGlnArgGly---LeuGlnTyrLeuValGlyMetLeuArgGl   543
            |:|     |:|   ||:||:
Db   2337 GTTGAGGAAGTGTCAGCTGGAGTCCGGAGTGCTCCGCAGGAGAGTGCTTCGTGCCGGCAC  2396

Qy    543 nSerProAlaLeuThrThrLeuAspAspArgGlyGlyCysGlnLeuProAlaProMetValTh   563
            ::|||   ||:||:
Db   2397 CAACCCACATCTCGTTGAGCTGGACCTGAGCGAAATGCACTGGAGGATTTGGGCCTGAG  2456

Qy    563 rTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuThrLeu---------  579
            ||:|
Db   2457 GTTACTATGCCAGGACTGAGGCACCCCAGTCTGCAGACTACGGACTTTGTGGCTGAAGAT  2516

Qy    580 ------SerLeuAlaSerValGluLeuSerGl::GnSe  590
            |||
Db   2517 CTGCGCGCCCTCACGTCCTGCTGCTGTGACCAGTCGCCTCAACTCTCAGTGTGAACCAGAG  2576

Qy    590 rLeuGlnGluLeuGlnAlaValLysArgAlaLysProAspLeuValIle----------  606
            |:|   :|||
Db   2577 CCTGAGAGAGCTGGACCTGAGCCTGAATGAGCCTGGGGAGCTCCGGGGTGCTGCTGCTGTG  2636

Qy    607 ---ThrHisPro-----  609
            |||:|||
Db   2637 TGAGGGCCTCAGGCATCCCATGTGCAAGTCCCAGACCCTGCCAAGTCCAGGTTGGCCATCTGCCGGCGCT  2696

Qy    610 -----------------AlaLeuAspGlyHisProGlnPrpProPro  618
            ||| ||||||:   ||||||:
Db   2697 GGGCTCTGCGCCGGGCTCTGCGCCTGTCTTTCTGTGGTCTGAGGGTCTTTCGTGGTCTCCA-GGCCAACCAACCTTCCG  2753
```

```
Alignment Scores:
Pred. No.:                          4.68e-84         Length:         3186
```

Db    2778  CTCTGCCGCCTGTGAGGGTCTTCTGTGGTGTCTCCA-GGCCAACCACCAACCTCCG 2831

RESULT 4
US-10-132-967-6
; Sequence 6, Application US/10132967
; Publication No. US20030170841A1
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THI
;   FILE REFERENCE: 07334-311002
;   CURRENT APPLICATION NUMBER: US/10/132,967
;   CURRENT FILING DATE: 2002-04-26
;   PRIOR APPLICATION NUMBER: US/09/653,901
;   PRIOR FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: US 09/506,067
;   PRIOR FILING DATE: 2000-02-17
;   NUMBER OF SEQ ID NOS: 13
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
;   LENGTH: 3102
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-132-967-6

Alignment Scores:
Pred. No.:              3.35e-81
Score:                  800.50
Percent Similarity:     43.64%
Best Local Similarity:  28.44%
Query Match:            24.88%
DB:

| | Length: | 3102 |
| | Matches: | 229 |
| | Conservative: | 117 |
| | Mismatches: | 241 |
| | Indels: | 193 |
| | Gaps: | 17 |

US-10-029-347-2 (1-625) x US-10-132-967-6 (1-3102)

Qy     1  MetLeuAlaGlnProGlnProLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
Db   862  ATCGTGAGAAAACCCTCAGATCCTTCCTCCAGATCCCTGATGAGCTGCAAGGT  921

Qy    21  LeuGlyGlyProGluAlaAlaPro---CysThrAspProPheGluAlaAlaSerGlyAla  39
Db   922  GCCTTTGACGAGCACATAGGACCGGCTTCGCACTGACTGGCAGAAGGCCGAGCGGGGGAC  981

Qy    40  ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThr  59
Db   982  ATTCTCCTGAGCAGCCCTCATCAGAAAGAAGACTGCTTCCCGAGGCCTCGTCCTCATCACC 1041

Qy    60  ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu  79
Db  1042  ACGAGACCTGGGCCCTGGAGAAACTGCAGCACTTGCTGGACCATCCTCGGCATGTGGAG 1101

Qy    80  ValArgGlyPheSerAspLysAspProIleLysTyrPheTyrLysPhePheArgAspGlu  99
Db  1102  ATCCTGGGTTTCTTCCGAGGCCAAAGGAAGAGTACTTTCTTCAGTACTACTTCTGTGATGAG 1161

Qy   100  ArgArgAlaGluArgGlyPheArgPheValLysGluAsnGluThrLeuPheAlaLeuCys  119
Db  1162  GCCCAAGCCAGGGCAGCCTCGATTCAGTCGAATCAGGAGAACGAGGTCCTCTTCACCATGTGC 1221

Qy   120  PheValProPheValCysTrpIleValCysThrValLeuThrValLeuArgGlnLeuGlu  139
Db  1222  TTCATCCCCGGTGCTGCTGGATCGTGTGCACTGGTGCACTGGAACACAGAGAGTGGC 1281

Qy   140  ArgAspLeuSerArgArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSer  159
Db  1282  AAGAGCTTGCCAGCACATCTAAGACACTCTTCCGCGTGTACGTCTTCTTCCCAGT 1341

Qy   160  ValLeuSerSerAlaProValAlaAspGlyProArgLeuGlyProArgLeuAspLeuArgAsnLeu  179
Db  1342  TTGCTGCAGCCCCGGGAGGGAGCCCAGGAGCCAGAGCCAGCCACCTCTGGGGGGCTC 1401

Qy   180  CysArgLeuAlaArgGluGlyGlyValLeuGlyValArgArgAlaGlnPheAlaGluLysGluLeu  199
Db  1402  TGCTCTTTGGCTGCGAGCAGATGGAAATCTGGAACCAGAAAAATCCTGTTTGAGGAGTCCGACCTC 1461

---

Db   1777  --------------------------:::|||:::|||          :::  |||  :18CC
Qy          CAGTGGATCCAAAGCAAAGCTCAG-----

Qy   331  AlaLysGlyLeuGluAspThrGluProGluGluGlyLysGlyGluGlnProAsn  350
Db  1801  --------------------------AGCGACGGCTCCACCCTGCGACAGGGCTCC------  1830

Qy   351  TyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluGlnAlaPheValArgGln  370
Db  1831  -----TTGGAGTCTTCAGCTGCTTGTACGAGATCCAGGAGGAGGAGTTTATCCAGCAG  1884

Qy   371  AlaLeuCysArgPheProGluAlaLeuAsnLeuGlnArgValArgPheCysArgMetAspVal  390
Db  1885  GCCCTGAGCCACTTCCAGGTGATCGTGGTCAGCAGTGTCAGCAACATTGCC---TCCAAGATGGAGCAC  1941

Qy   391  AlaValLeuSerTyrCysVal.ArgCysProAlaGlyAlaGlyGlnAlaLeuArgLeuIle---  409
Db  1942  ATGGTCTCCTCGTCTGTCGAAGCGCTGCAGGAGCGCCAGGTGCTGCA-CTTGTATGG  2000

Qy   410  --------------------------------------------SerCysArgLeu  413
Db  2001  CGCCACCTACAGCGCCGGACGCGGGAAGACCGGCGGAGGTGCTCCGCAGGAGCGCACACGCT  2060

Qy   414  ValAlaAlaAlaGlnGluLysLeuLysLysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGly  433
Db  2061  GTTGGTGCGACGAGCTACCAGAGAGACCGGTTCTGCTGA---CGCCTACAGTGAACATCTGGC  2117

Qy   434  GlyGly---------SerSerGlnGlyThrThrLysGlnLeuProAla-SerLeu-  448
Db  2118  AGCGGCCCTGTGCACCAATCCAAACCTGATAGAGAGCTGTCTCTGTACCGCAAAATGCCCTGGG  2177

Qy   449  -------LeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSer  465
Db  2178  CAGCCGGGGGGTGAAGCTCTCTGTCAGGACTCAGACACCCCAACTGCAAACTTCAGAA  2237

Qy   465  rLeuThrLeuSerHisCysLysLysLeuProAspAlaValCysLysSerArgAspLeuSerGluAlaLe  485
Db  2238  CCTGAGGGTGAACAGAGGACCATCTCAGCTCAGGCTGCGGAGGACTCCTGTGCAGCTCT  2297

Qy   485  tArgAlaAlaAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGluAlaG-  505
Db  2298  CATAGCCAATAAGAATTTGACAAGGATGATCTCAGTGCAACGGCGTTGGATTCCCAGG  2357

Qy   505  yLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrValArgVa  525
Db  2358  CATGATGTGCTTTGCGAGGGCCCTGCGGCATCCCCAATGCAGGCTGCAGATGATTCAGTT  2417

Qy   525  GlnLeuProAspProGlnArgGly--LeuGlnTyrLeuValGlyMetLeuArgGlnSe  544
Db  2418  GAGGAAGTGCAGCTGGAGTCCGGGGCTTGTCAGGAGATGGCTTCTGTGCTTGGCACCAA  2477

Qy   544  rProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProProMetValThrTy  564
Db  2478  CCCACATCTGGTTGAGTTGAGTTGGCAACGGCGTTGGGCCTGGAGG  2537

Qy   564  rLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeu---------  579
Db  2538  ACTATGCCAGGGACTGAGGCACCCAGTCTGCAGACTACGGACTTTGTGCCTGAAGATCTG  2597

Qy   580  ------------------SerLeuAlaSerValGluLeuSerGluGlnSerLe  591
Db  2598  CCGGCTCACTGCTGCCTGTGACGAGCTGGCCTCAACTCCAGTGGAACCAGGAGCCT  2657

Qy   591  uGlnGluLeuGlnAlaValLysArgAlaLysProAspLeuValIle---------  606
Db  2658  GAGAGAGCTTGACCTGAGCCTGAATGAGCTGGGGGGACCTCGGGGTGCTGCCTGCGTGTGA  2727

Qy   607  ------ThrHisPro-----------------  609
Db  2727  GAGAGCTTGACCTGAGCCATCCCAACGTGCGGGTTGGGCATCCGCCGGCTGGG

Qy   610  --------------AlaLeuAspGlyHisProGlnProPro  618
Db  1402  TGCTCTTTGGCTGCGAGCAGATGGAATCTGGAACCAGAAAAATCCTGTTTGAGGAGTCCGACCTC 1461

```
Qy   460  --------LeuCysHisLeuSerSerLeuThrLeuSerHisCysLysLysProAspAla  476
                  |||||||:  ||  |:::||||:|||||  |||:||||||||||||||
Db  2461  AAGCACCTGTTGTGCAATCTGAAGAGAGCTCTGGTTGTCAGCTGCTGCTCACATCAGCA  2520

Qy   477  ValCysArgAspLeuSerGluAlaLeuArgAlaAlaProAlaLeuThrGluGlyLeu   496
                 :::  ||||||||:::::|||||  |||:||||||||
Db  2521  TGTTGTCAGGATCTTGCACTAGTATTGAGCACCAGCCATTCCTGACCGAGACTCTATGTG  2580

Qy   497  LeuHisAsnArgLeuSerGluAlaGlyAlaGlyMetLeuSerGluGlyLeuAlaTrpPro  516
                 ||||:::||  |||||||||||:||||  |||:||||||||:
Db  2581  GGGGAGAATGCCTTGGGAGACTCAGGAGTCGCAATTTTATGTGAAAAAGCCAAGAATCCA  2640

Qy   517  GlnCysArgValGlnThrValArgVal----------   525
                 :|||||||||  |  ||:::
Db  2641  CAGTGTAACCTGCAGAAAACTGGGGTTGGTGAATTCTGGCCTTACGTCAGTCTGTTGTTCA  2700

Qy   525  ----------------------------------------   525
Db  2701  GCTTTGTCCTCGCTACTGACACTAATCAGAATCTCACGCACCTTTACCTGCGAGGCAAC  2760

Qy   526  GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuVal--GlyMetLeuArgGlnSer  544
                 |:::::||  |:::  |||:::|||||||:||||  ||
Db  2761  ACTCCGGAGAC---------AAGGGATCAAACTACTCTGTGAGGGACTCTTGCACCCCGAC  2814

Qy   545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMet----------  561
                 ||:::||||||||:||  |:::  ||||||||
Db  2815  TGCAAGCTTCAGGTGTTGGAATTAGACAACTGCAACCTGCACCTCAGTCAGCTGCTGGAT  2874

Qy   561  ----------------------------------------   561
Db  2875  CTTTCCACACTTCGACCTCCAGCCAGCCTGCGAAAGTCAGCCTGGGCAACAATGAC  2934

Qy   562  --------ValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLysLeu  576
                 |||:||||||||||||:::|||||||||:||:
Db  2935  CTGGGCGACCTGGGGGTCATGATGTTCTGTGAAGTGCTGAAACAGAGAGCTGCCTCCTG  2994

Qy   577  GlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGlnLeuGlnAla  596
                 ||  ||||||:::||  |:::  |:::|||||||||||
Db  2995  CAGAACCTGGGGTTGTCTGAAATGTATTCAATTATGAGACAACAAAAGTGCGTTAGAAACA  3054

Qy   597  ValLysArgAlaLysProAspLeuValIle  606
                 :|||||:::|||:||||
Db  3055  CTTCAAGAGAGAAAAGCCTGACCGTGACCGTC  3084
```

```
Qy   200  GluGlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLysGluLeu  219
                 |||:  :::  |||  ||||:  |||:||
Db  1462  AGGAATCATGGACTGCAGAGAGGCGGATGTG---TCTGCTTCCTGAGGATGAACCTGTTC  1518

Qy   220  ProGlyValLeuGluThrGluThrGlyIleValThrTyrGlnPheIleAspPheGlnGluPhe  239
                 ::||||||||:||||||||||||||||||
Db  1519  CAAAAGGAAGTGGACTGCGGAGAAGTTCTACAGCTTCATCCACATGACTTTCCAGGAGTTC  1578

Qy   240  LeuAlaAlaLeuSerTyrLeuGluAspGlyLeuValProArgThrAlaAlaGlyGly  259
                 ||||||||||:::||||||||||||
Db  1579  TTTGCCGCCATGTACTACCTTGCCGAAGAGAAAAGGAAGGAAGGACGAACGTTCCAGGG  1638

Qy   260  ----------ValGlyThrLeuLeuArgGlyAspAlaGln---  269
                 ||:::
Db  1639  AGTCGTTGAAGCTTCCAGCCGAGACGTGACAGTCCTTGTCLeuSerLeuSerAlaGluArg  1668

Qy   270  ProHisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuSerLeuAlaGluArg  289
                 ::||||||:  |||||:||||||||||
Db  1699  GAAAAGGGGTATTTGATTTTGTTGTACGTTTCCTCTTGGCCTGGTAAACCAGGAGGAGG  1758

Qy   290  MetArgAspGlyArgHisPheGlyCysMetValSerValAlaLysValLysGlnAla  309
                 :::|||||||||:::||||||
Db  1759  ACCTCCTACTTGGAAGAAGAAATTAAGTGCAAGATCTCTCAGCAAAATCAGGCTGGAGCTG  1818

Qy   310  LeuArgTrpValGlnGlyGlnGlyGlnGlyValAlaProGlyValValAlaThrGlu  329
                 :::||||:::
Db  1819  CTGAAAATGGATT------------------------GAAGTGAAAGCC  1842

Qy   330  GlyAlaLysGlyLeuGluAspThrGluProGluGluProGluGluGluGlyGlyGluPro  349
                 |||  :::
Db  1843  AAAGCTAAAAAGTGCAG-----ATCCAGCCCAGCCAG----------  1875

Qy   350  AsnTyrProLeuGluLeuLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheValArg  369
                 :||||||||||:||:||||||||
Db  1876  ----CTGGAATTGTTCTACTGTTTGTACGATGAGGAGGACTTCGTGCAA  1926

Qy   370  GlnAlaLeuCysArgPheProGluLeuAlaLeuGlnAlaLeuGlnArgValArgPheCysArgMetAsp  389
                 |||  :::  |||:::
Db  1927  AGGGGCATGGACTATTTCCCCAAGATTGAGATCAAT------CTCTCCACCAGAATGGAC  1980

Qy   390  ValAlaValLeuSerTyrCysValArgCysCysProAlaGlyLeuGlnAlaLeuArgLeuIle  409
                 |||
Db  1981  CACATGGTTCTTCTTCCTTTGCATTGAGAACTGTCATCGGGTGAGTCACTGTCCCTGGGG  2040

Qy   410  SerCysArgLeuValAlaAlaGlnGluLysLysLysSerLeuArgGlyGln  429
                 ::  ||||
Db  2041  TTTCTCCATAACAACATGCCCAAGAGGAGAAGAAAGAAAGAAAGCCGACACCTTGAT  2100

Qy   441  --------AlaSerLeuGlyGlySerGlnGly--ThrThr  441
                 ||  |||
Db  2101  ATGGTGCAGTGCTGCCTCCCAGCTCCTCTCATGCTCCTGTTCTCATGGATTGGTGAAC  2160

Qy   442  LysGlnLeuProAlaSerLeuLeuHisProLeuPhe------------GlnAla  455
                 :::||||  |||
Db  2161  AGCCACCTCACTTCCAGTTTTGCGCGGGCCTCTTTCAGTTCTGAGCACCAGCCAGAGT  2220

Qy   456  MetThr------  459
                 ::|||
Db  2221  CTAACTGAATTGGACCTCAGTGACAATTCCTCTGGGGGACCCAGGGATGAGGAGAGTGTTGTG  2280

Qy   459  ------------------------  459
Db  2281  GAAAACGCTCCAGCATCCTGGCTGTAACATTCGGAGATTGTGGCTTGGGCGGCCTGTCGCCTC  2340

Qy   459  ------------------------  459
Db  2341  TCGCATGAGTGCTGCTTCGAATCCCTTGCATCCCTTGGTCCTCAGCAGCAACCAGAAGCTGGTGGAG  2400

Qy   459  ------------------------  459
Db  2401  CTGGACCTGAGTGACAACGCCCTCGGTGACTTCGGAATCAGAGACTTCTGTGTGGTGGGGACTG  2460
```

```
Pred. No.:
Score:                   3.35e-91
Percent Similarity:      800.50        Length:      3102
Best Local Similarity:   43.64%        Matches:     219
Query Match:             28.44%        Conservative:117
DB:                      24.88%        Mismatches:  241
                                       Indels:      193
                                       Gaps:        17
```

US-10-029-347-2 (1-625) x US-10-127-516-6 (1-3102)

```
Qy    1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuPro       20
          :::  ::  |||   :::|||||||||||||:::|||
Db  862 ATCGTGAGAAAACCCTCGAGATCCTTCCTCATGGACGGCTTCGATGAGCTGCAAGGT      921

Qy   21 LeuGlyGlyProGluAlaPro--CysThrAspProPheGluAlaSerGlyAla        39
          :::|||  ::  |||                 |||||||||||||::::
Db  922 GCCTTTGAGCAGCACATAGGACCGGTCTGCACTGACTGGCAAGGCCGAGCGGGGAGAC      981

Qy   40 ArgValLeuGlyLeuLeuSerLysAlaLeuLeuProThrArgAlaLeuLeuLeuValThr        59
          :::|:|  |||  :::|||||||||||||:::::
Db  982 ATTCTCCTGAGCAGCCTCATAGAAAGGAGCTGCTTCCCGAGGCCTCTCTGCTCATCACC     1041

Qy   60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyLeuArgLeuCysSerProGlnCysAlaGlu        79
          |||  :::|||||||||||||  :::
Db 1042 ACGAGACCCGTGCCCTGGAGAAACTGCACTGCTGGCATCCTCGGCCATGTGGAG      1101

Qy   80 ValArgGlyPheSerAspLysAspLysTyrLysTyrLeuLysTyrLysPhePheArgAspGlu        99
          ::::::|||
Db 1102 ATCCTGGGTTCTCCGAGGCCAAAGGAAGAGTACTTCTTCAAGTACTTCTTCTGATGAG     1161

Qy  100 ArgArgAlaAlaGluArgArgAlaTyrArgPheValLysGluThrGlnLysGluThrGlnLysGlu       119
          |||  :::::::
Db 1162 GCCCAAGTCCAGGCAGCCTCAGTCCTGATTCAGGAGAACGAGGTCCTTCTTCACCATGTGC     1221

Qy  120 PheValProPheValCysTrpIleValCysThrValGlnLeuArgGlnLeuLeuGly       139
          |||:::|||||||||||
Db 1222 TTCATCCCCTGTCTGCTGGATCGTGTGGTGCACTGCACAGCAGATGGAGTGGGC      1281

Qy  140 ArgAspLeuSerArgThrArgSerThrThrThrTyrLeuLeuPheIleThrSer       159
          :::|||
Db 1282 AAGAGAGCCTTGGCCAGACATCTAAGACCACCACGCGGTACGTCTTTCTTCCTTTCCAGT     1341

Qy  160 ValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnLysTyrAspLeuArgAsnLeu       179
          :::|||
Db 1342 TTGCTGCAGCCCCGGGAGGGAGCCCAGGAGACACGGCCTCTGCGCCCACCTCTGGGGGTC     1401

Qy  180 CysArgLeuAlaArgGluGluValValLeuGlyValArgArgAlaGlnAlaGlnLeuGluLeu       199
          |||  :::|||||||
Db 1402 TGCCGTTCTTGGCTGCAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCTG     1461

Qy  200 GluGlnLeuGlnLeuGluLeuArgGlyLeuSerValLysGlnThrLeuLysSerLysLeu       219
          |||||||||||||
Db 1462 AGGAATCATGGACTGCAGAAGTCGCGGATGTG----CTGCTTCCTGAGGATGAACCTGTTC     1518

Qy  220 ProGlyValLeuGluThrGluValThrThrTyrGlnPheIleLeuAspSerGlnSerPhe       239
          :::|||
Db 1519 CAAAAGGAAGTGGACTGGCAGAGTTCTACAGCTTCCACCATGACTTCCAGGAGTC     1578

Qy  240 LeuAlaAlaLeuSerTyrLeuLeuGluGluArgThrAlaAlaG:yG:y       259
          :::|||  :::|||||||
Db 1579 TTTGCCGCCATGACTACCTGCTGCTGGAGAAGAAGCGAGGAAGCGAACGTTCCAGGG     1638

Qy  260 ------------ValGlyThrLeuArgGluArgGlyAlaGln--       269
                             :::||||||
Db 1639 AGTCGGTTTGAAGCTTCCAGCTTGAAGCTGAGACGTGACAGTCCTTCTGAAACTATGTGGGAAAACTATGCCAAATTC     1698

Qy  270 ProHisSerHisValLeuThrHisArgPheLeuPheLeuTyrLeuLeuSerAlaLeuArg       289
          |||  :::
Db 1699 GAAAAGGGTATTTGATTTTGTTGTACGTTTCCTCTTTTGGCCTGGTAAACCAGGAGAGAG     1759

Qy  290 MetArgAspIleGluArgHisPheGlyCysMetValSerGlyValLysValGlnGluAla       309
          :::|||::::::
Db 1759 ACCTCCTACTTGAGAAGAAATTAAGTTGCAGAATCTTGCAAATCAGGCTGGAGCTG     1818
```

```
Qy  310 LeuArgTrpValGlnGlyGlnGlyCysProGlyValAlaProGluValThrGlu       329
          |||:::
Db 1819 CTGAAATGGATT--------------------------------GAAGTGAAAGCC     1842

Qy  330 GlyAlaLysGlyLeuGluAspThrGluProGluGluGluGluGlyGlyGlyGluPro       349
          :::|||      :::
Db 1843 AAAGCTAAAAAGCTGCAG---------ATCCAGCCCAGCCAG-----     1875

Qy  350 AsnTyrProLeuGluLeuLeuTyrCysLeuTyrGluThrGlnGlnAspAlaPheValArg       369
          ------CTGGAATTGTTCTACTGTTTGTACGAGTGCAGGAGGACTTCGTGCAA     1926

Qy  370 GlnAlaLeuCysArgPheProGluLeuAlaLeuGlnLeuArgValAlaArgPheCysArgMetAsp       389
          :::||||||
Db 1927 AGGGGTCCATGACTATTCCCCAAGATTGAGATCAAT-----CTCTCCACCAGAATGGAC     1980

Qy  390 ValAlaValLeuSerTyrCysValArgCysCysProAlaGlyValGlnAlaLeuArgLeuIle       409
          |||:::
Db 1981 CACATGTGTTTCTCTTTGCATTGAGAACTGTCATCGTGGTGCTCACTGTCCCTGGGG     2040

Qy  410 SerCysArgGlyLeuValAlaAlaAlaGlnGluLysLysLysSerLeuGlyLysLysArgLeuGln       429
          :::|||
Db 2041 TTTCCCATAACATGCCCAAGGAGGAAGGAGGAGAAAAGAAGGCCGACACCTTGAT     2100

Qy  430 --------------AlaSerLeuGlyGlyGlySerGlnGly---ThrThr       441
                                :::|||
Db 2101 ATGGTGAGTCGTGTCCTCCCAAGCTCCTCATGCTGCCTGTTCTCATGGAGTGGTGAAC     2160

Qy  442 LysGlnLeuProAlaSerLeuLeuHisProLeuPhe--------------GlnAla       455
          :::||||||
Db 2161 AGCCACCTCACTTCAGTTTTTGCCGGGGCCTCTTTCAGTTCTGAGCACCAGCCAGAGT     2220

Qy  456 MetThr----------------------------AspPro       459
          :::|||
Db 2221 CTAACTGAATTGACCTCAGTGACAATTCTCTGGGGGACCCAGGATGAGAGTGTTGT     2280

Qy  459 ---------------------------------------       459
Db 2281 GAAACGTCCCAGCATCTCGTGTAACATTCGGAGATTGTGGTGGGGCGGCTGGCCTC     2340

Qy  459 ---------------------------------------       459
Db 2341 TCGGATGAGTGCTGCTTCGTGACATCCTCGGTCCTCAGCAGCAACCAGAGCTGTGGAG     2400

Qy  459 ---------------------------------------       459
Db 2401 CTGGACCTGAGTGACAACGCCCTCGGTGACTTCGGAATCAGACTCGTGTGTGTGGGACTG     2460

Qy  460 --------------LeuCysHisLeuSerSerLeuThrLeuSerHis:CysLysLysProAspAla       476
          :::|||
Db 2461 AAGCACCCTGGTTGCAATCGGGTCTCGGTTGGTCAGCTGCTCGCTGCCTCACATCAGCA     2520

Qy  477 ValCysArgAspLeuSerGluAlaLeuArgAlaAlaAlaProAlaLeuThrGluLeuGlyLeu       496
          |||
Db 2521 TGTTGTCAGGATCTTGCATCAGTATTGAGCACCACCAGCCATTCCCTGACCAGACTCTATGTG     2580

Qy  497 LeuHisAsnArgLeuSerGluAlaGlyLeuAlaArgMetLeuSerGluGluLeuAlaTrpPro       516
          |||:::
Db 2581 GGGGAGAATGCCTTGGGAGACTCAGGAGTCGCAATTTATGTGAAAAAGCCAAGAATCCA     2640

Qy  517 GlnCysArgValGlnThrValArgVal---       525
          :::|||
Db 2641 CAGTGTAACCTGCGAGAAACTGGGGTTGGTGAATTCTGGCTTACGTGTTGTTCA     2700

Qy  525 ------------------------------       525
Db 2701 GCTTTGTCCTCGGTACTCAGCACTAATCAGATTCAGAATCCACGCACCTTACCTGCGAGGCAAC     2760

Qy  526 GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuGlyMetLeuArgGlnSer       544
          :::|||                    --GlyMetLeuArgGlnSer
Db 2761 ACTCCGGAGAC--------AAGGGGATCAAACTACTCCTGTGAGGGGACTCTTGCACCCGAC     2814

Qy  545 ProAlaLeuThrLeuAspLeuSerGlySerGlyGlnLeuLeuProAlaProMet---------       561
```

```
Db  1102 ATCCTGGGTTCTCCGAGGCCAAAAGGAAAGAGTACTCTTCAAGTACTTCTCAAGTGATGAG 1161
          |||                              |||            ::|
Qy   561 TGCAAGCTTCAGGTGTTGGAATTAGAGAACTGCAACCTCACGTCACACTGCTGCCTGGGAT 2874

Qy   100 ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys   119
          |||                                                    :|:
Qy   561                                                              561

Db  1162 GCCCAAGCCAGGGCAGCTTCAGTCTGATTCAGGAGAACGAGGTCCTTCACCATGTGC    1221
          |||||        |||                         |||      |||
Db  2875 CTTTCCACACTTCTGACCTCCAGCCAGAGCCCTGGAAAGCTGAGCCTGGCAACAATGAC   2934

Qy   120 PheValProPheValCysTrpIleValCysThrValLeuArgIlnGlnLeuGluLeuGly  139
          |||                                    |||::::|||     |||
Qy   562         ------ValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeu   576

Db  1222 TTCATCCCCTGGTCTGCTGATCGTGTGCACTGACTGAAACAGCAGATGGAGAGTGGC    1281
          |||:::                     |||              |||         |||
Db  2935 CTGGCGACCTGGGGTCATGATGTTCTGTGAAACAGCAGAAACAGCCTGCCTCCTG   2994

Qy   140 ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSer    159
          :::     |||      |||::      |||:::       |||:::       |||
Qy   577 GlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGlnGluGluGlnAla  596

Db  1282 AAGAGCCTTGCCAGACATCTTAGAGACATTCAATTATGAGACAAAAAGTGCGTTAGAAACA 1341
          |||           |||              |||:::
Db  2995 CAGAAACCTGGGGTTGTCTGAAATGCATTCAATTATGAGACAAAAAGTGCGTTAGAAACA   3054

Qy   160 ValLeuSerSerAlaProValAlaAspIlyProArgLeuLeuGlnGlyAspLeuArgAsnLeu 179
          :::||||
Qy   597 ValLysArgAlaLysProAspLeuValIle   606

Db  1342 TTGCTGCAGCCCCGGGAGGGGAGCCAGGAGCACGGCCTCTGCGCCCACCTCTGGGGGCTC  1401
          ::|||
Db  3055 CTTCAAGAGAGAAAAGCCTGAGCTGACCGTC   3084

Qy   180 CysArgLeuAlaArgGluGlyValLeuLeuGlyValArgValArgAlaGlnPheAlaGluLysGluLeu 199
          ::|||    |||
Db  1402 TGCTCTTTGGCTGCAGATGGAATCTGCAGCAAACAGAAAAATCCTGTTTGAGGAGTCCGACCTC   1461

Qy   200 GluGlnLeuGluLeuGlySerLysValGlnThrLeuPheLeuSerLysLysGluLeu     219
          |||:::                    |||:::
Db  1462 AGGGATCATGACCGCAGAAGGGGATGTG----TCTGCTTCCTGAGAATGAACCTGTTC  1518

Qy   220 ProGlyValLeuGluThrGluValThrTyrGlnPheIleLeuAspAlaSerGlnSerPheGlnGluPhe 239
          |||||||||                     |||:::
Db  1519 CAAAAGGAAGTGGACTGCGAGAAGTTCATCCACATGACTTTCCAGGAGTTC  1578

Qy   240 LeuAlaAlaLeuSerTyrLeuLeuGluAspGlyGlyGlyValProArgThrAlaAlaGlyGly 259
          |||||||||
Db  1579 TTTCCCGGACCATGTACTACCTGTGGAAGAGAAAAGGAAGGAAGGACGAAGTTCCAGGG  1638

Qy   260          ------ValGlyThrLeuGluArgGlyAspAlaGln--     269
                         :::
Db  1639 AGTCGTTTGAAGCTTCCCAGCCGAGACGTGACAGTCCTTCTGAAAACATGGCAAATTC    1695

Qy   270 ProHisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerLeuSerAlaGluArg 289
          :::
Db  1699 GAAAAGGGGTATTTGATTTTGGTTGTACGTTTCCTTTGGCCTGGTAAACAGGAGAGG  1758

Qy   290 MetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluAla 309
          ::||||
Db  1759 ACCTCCACTTGGAGAAGAAGAAATTAAGTTGCAAGATCTCAGCAAATCAGCCTGGAGCTG   1818

Qy   310 LeuArgTrpValGlnGlyGlnGlyCysProGlyValAlaProGluValThrGluGlu     329
          |||||||::
Db  1819 CTGAAATGATT------------GAAGTGAAAGCC   1842

Qy   330 GlyAlaLysGlyLeuAspThrGluProGluGluGluGluGluGluGluPro     349
          ::||||
Db  1843 AAAGCTAAAAAAGCTGCAG------ATCCAGCCCAGCCAG-----    1875

Qy   350 AsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheValArg     369
          |||||||                            ||||||
Db  1876 ---------CTGGAATTGTTCACTGTTTGTACGAGATGATGCAGGAGGAGGACTTCGTGCAA   1926

Qy   370 GlnAlaLeuCysArgPheProGluLeuAlaGlnLeuArgValArgPheCysArgMetAsp 389
          |||:::                           ----
Db  1927 AGGGCCCATGGACTATTCCCCAAGATTGAGATCAAT------CTCTCCACCAGAATGGAC   1980

Qy   390 ValAlaValLeuSerTyrCysValArgGlyCysProAlaGlyValGlnAlaLeuArgLeuIle 409
          ::|||
Db  1981 CACATGGTTTCTCCTTTGCATTGAGAACTGTCATCGGGTGAGTCACTGTCCCTGGGG  2040

Qy   410 SerCysArgLeuValAlaAlaGlnLeuGluLysLysLysSerLeuLysLysArgLeuGln   429
          :::                           ::|||
Db  2041 TTTCTCCATAACATGCCCAAGGAGGAGAAGAAGGACGAAGGCCGAACACCTTGAT   2100

Qy   430          ------AlaSerLeuGlyLysGlySerGlnGly---ThrThr 441
                         ::|||
Db  2101 ATGTGCAGTGTGTCCTCCCAAGCTCCTCATGCTGCCTGTCTTCTCATGGATTGGTGAAC  2160
```

Alignment Scores:
Pred. No.:
Score: 800.50
Percent Similarity: 43.64%
Best Local Similarity: 28.44%
Query Match: 24.88%
Db:

3.5e-81

Length: 3102
Matches: 219
Conservative: 117
Mismatches: 241
Indels: 193
Gaps: 17

US-10-029-347-2 (1-625) x US-10-027-629-6 (1-3102)

```
Qy    1 MetLeuAlaGlnProGlnLeuLeuGlnPheIleLeuLeuAspGlyAlaAspGluLeuPro Ala 20
          |||:::
Db  862 ATCGTGGAGAAAAACCCCCAGATCTCTTCCTCATGGACGGCTTCGATGAGCTGCAAGGT  921

Qy   21 LeuGlyGlyProGluAlaAlaPro--CysThrAspProPheGlualaAlaSerGlyAla   39
          :::||||
Db  922 GCCTTTGACGAGCACATAGGACGCGCTCCGACTGGACGAAGGCCGAGCGGGAGAC   981

Qy   40 ArgValLeuGlyLeuGlyLeuLeuSerLysLysSerLysAlaLeuProThrAlaLeuLeuLeuValThr 59
          ::|||
Db  982 ATTCTCCTGACGCAGCCTCATCAGAAAGAAGAGCTGCTTCCCGAGGCCTCTGCTCATCACC  1041

Qy   60 ThrArgAlaAlaProGlyArgGluGlnGlyLeuLeuArgLeuCysSerProGlnCysAlaGlu  79
          :::|||
Db  1042 ACGAGACCTGTGCCCTGGAGAAACTGCAGACTCTTGCTGACCATCCTCGGCATGTGGAG  1101

Qy   80 ValArgGlyPheSerAspLysLysAspLysLysTyrPheTyrLysLysAspGluGlu  99
          :::
Db  1102 ...
```

```
; CURRENT APPLICATION NUMBER: US/10/132,967
; CURRENT FILING DATE:      2002-04-26
; PRIOR APPLICATION NUMBER: US/09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; SOFTWARE: FastSEQ for Windows Version 4.0
; NUMBER OF SEQ ID NOS: 13
; SEQ ID NO 4
;   LENGTH: 3857
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-132-967-4
```

```
Alignment Scores:
Pred. No.:             4.66e-81        Length:          3857
Score:                 800.50          Matches:         219
Percent Similarity:    43.64%          Conservative:    117
Best Local Similarity: 28.44%          Mismatches:      241
Query Match:           24.88%          Indels:          193
DB:                    12              Gaps:            17
```

US-10-029-347-2 (1-625) x US-10-132-967-4 (1-3857)

```
Qy     1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla     20
           ::::::  :::::||    |::::: ||||| ||||||||:::||||||:  ||:||||||
Db  1000 ATCGTGAGAAAACCTCCAGAATCCTCTTCCTACGAGGCTTCGATGAGCTGCAAGGT      1059

Qy    21 LeuGlyGlyProGluAlaAlaPro---CysThrAspProPheGluAlaAlaSerGlyAla    39
           :|||:::|||||:||||||||        :::: :|||||::|||||||||||||||||
Db  1060 GCCTTGACGAGCACATAGGACCGCTCGCACTGAC·GGCAGAAGGCCCGAGCGGCGGGAGAC  1119

Qy    40 ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThr    59
           :::| |||:|||||:|||:|||||||||::|||||||:|||||||||||::|:|||||
Db  1120 AATTCTCCTGACGAGCACGTCATCAGAAAGAAGAGCTGCTTCCCGAGGCCTCTCGCTCATCACC   1179

Qy    60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu    79
           ||||  :::||||||||:||||||||||||:||:||:|||||||||||||||||||||||
Db  1180 ACGAGACCTGTGGCCCTGGAGAAACTGCAGCACTGCTGCGACCATCCTGGCATGTGGAG   1239

Qy    80 ValArgGlyPheSerAspLysAspLysLysTyrPhePheLysAspGlu    99
           :::  :::|||::|||||||||:||||||||||||||||:||:|||:::
Db  1240 ATCCTGGGTTTCTCCGAGGCCAAAAGGAAGAGTACTTCTTCAAGTACTTCTCTGATGAG   1299

Qy   100 ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys   119
           ::::::::::::::::::|:::|||:::|||:::|:|||||||||||||||||||
Db  1300 GCCCAGACCCAGGGACCGCAGCCTCCAGTCTGATTCAGGAGAAGCGAGGTCCTCTTCACCATGTGC   1359

Qy   120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGlnLeuGly   139
           :::::::::::::: :::::: :::|::::::::::::::::||::|::::::
Db  1360 TTCATCCCCCTGCTCGCTGCTGGTGTGACTGTGTGACTGCTGAAACAGCAGA·GAGAGTGGC   1419

Qy   140 ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuPheLleThrSer   159
           :::  :::::::::::|::|||::::::::::::|::|::|||||||||||||
Db  1420 AAGAGACCTTGCCCAGACCATCCAAGACCTAAGACCCGGGTACGTCTTCTCCTTTCCAGT   1479

Qy   160 ValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeu   179
           :::::||:                       |||::|  ||||||||||||||||
Db  1480 TTGCTGCCCAGGGAGGGGAGCCCAGCAACCACGGGCCTCGCGCCACCTGGGGGCTC   1539

Qy   180 CysArgLeuAlaArgGlyValLeuGlyValArgArgAlaGluAsnThrLeuLysGluLeu   199
           :::::::::::::::::::::::::::::::::::::::::::::::
Db  1540 TGCTCTTTGGCTGCAGATGGAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCTC   1599

Qy   200 GluGlnLeuGluLeuArgGlySerLysSerValGlnThrLeuPheLeuSerLysLysGluLeu   219
           :::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db  1600 AGGAATCATGGACTGCAGAAGGCGGATGTG---TCTGCTTTCCTGAGGATGAACCTGTTC   1656

Qy   220 ProGlyValLeuGluThrGluValThrTyrGlnPheIleAspPheGlnGluPhe   239
           :::: :::|::::||::|    :::::::::|::|||||||||||||||||||
Db  1657 CAAAAGGAAGTGGACTGCGAGAAGTTCTACAGCTTCATCCACATGACTTTCCAGAGTTC      1716

Qy   240 LeuAlaAlaLeuSerTyrLeuLeuGluLeuAspProGlyValProArgThrAlaAlaGlyGly   259
```

```
Qy   442 LysGlnLeuProAlaSerLeuLeuHisProLeuPhe------------GlnAla   455
           |||                                       :::::
Db  2161 AGCCACCTCACTTCCAGTTTTGCCGGGGGCCTCTTTGCCGAGGCACCAGCCGAGT      2220

Qy   456 MetThr------------------------------------AspPro-----------   459
           |||                                     :::
Db  2221 CTAACTGAATTGGACCTCAGTGACAATTCTCTGGGGGACCCAGGGATGAGAGTGTGTGT   2280

Qy   459 -----------------------------------------------------------  459
Db  2281 GAAAGCGCTCCAGCATCCTGGCTGTAACATTCGGAGATTGTGGTTGGGGCGGCTGTGGCCTC   2340

Qy   459 -----------------------------------------------------------  459
Db  2341 TCGGATGAGTGCTGCTTCGACATCTCCTTGGTCCTCAGCAGCAACCAGAAGCTGGTGGAG   2400

Qy   459 -----------------------------------------------------------  459
Db  2401 CTGGACCTGGAGTGACAACGCCCTCGGCGACTTCGGAATCAGAGACTTCGTGTGTGGGACTG   2460

Qy   460 -----LeuCysHisLeuSerSerLeuThrLeuSerHisCysLysLysLeuProAspAla   476
              |||:::|||||:::||||||:||||||:|:|||
Db  2461 AAGCACCTGGTTGTGTCAATCTGAAGAAGTCTCGGTTGGTGGTCAGTGCTGCCTCACATCAGCA   2520

Qy   477 ValCysArgAspLeuSerGlyAlaLeuArgAlaLeuAlaProAlaLeuThrGluLeuGlyLeu   496
           :::|||::::::||||||:::||||||||||||:::::::|||||:|::|:||||||:|||
Db  2521 TGTTGTCAGGATCTTGCATCAGTATTGAGCACCAGCCATTCCCTGACCAGACTCTATGTG   2580

Qy   497 LeuHisAspArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpPro   516
           ::|:::|||:|||||:::|||::::::|||::|||:::::::::|||||||
Db  2581 GGGGAGAATGCCTGGAGACTCAGGAGCTCGCAATTTTATGTGAAAAAGCCAAGAATCCA   2640

Qy   517 GlnCysArgValGlnThrValArgVal   525
           |||::|| :::::: :::
Db  2641 CAGTGTAACCTGCAGAAACTGGGGTTGGTGAATTCTGGCCTTACGTGCTGTCGT·GTTCA   2700

Qy   525 -----------------------------------------------------------  525
Db  2701 GCCTTGTCTCGTACTCAGCACTAATCAGAATCTCACGCACCTTTACCTGCGAGGCAAC   2760

Qy   526 GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuVal--GlyMetLeuArgGlnSer   544
           :::::: :::::::::::||||||:::||||::          |||  ::::::||
Db  276- ACTCTGGGAGAC------AAGGGATCAAACACTCGTGAGGGACTTTGCACCCCGAC      2814

Qy   545 ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnSerLeuProAlaProMet------   561
           ::::::::::::::::: ::::::|||::::: ::::
Db  2815 TGCAAGCTTCAGGTGTTGGAATTAGACAACTGCAACCTTCACGTCACACTGCTGCTGGAT   2874

Qy   561 -----------------------------------------------------------  561
Db  2875 CTTTCCCACACTTCTGACCTCCCAGCCAGGCCTGCGAAAGCTGAGCCTGGCAACAATGAC   2934

Qy   562 --------ValThrTyrLeuCysValAlaValLeuGlnHisGlnGlyCysGlyLeu   576
              :::::::::::::::::||::|:::|
Db  2935 CTGGGCGCCGTGGGGGTCATGATGTTCTGTGAAGTGCTGAAACAGCAGAGCTGCCTCCTG   2994

Qy   577 GlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGluGlnLeuGlyLeuGlnLeuAla   596
           |:::|||  ::::::::::::::::::::::
Db  2995 CAGAACCTGGGGTTGTCTGAAATGTATTTCAATTATGAGACAAAAAGTGCGTTAGAAACA   3054

Qy   597 ValLysArgAlaLysProAspLeuValIle   606
           :::::::::::::::::::::
Db  3055 CTTCAAGAGAAAAAGCCTGAGCTGACCGTC   3084
```

```
RESULT 7
US-10-132-967-4
; Sequence 4, Application US/10132967
; Publication No. US2003017084lAl
; GENERAL INFORMATION:
;    APPLICANT: Bertin, John.
;    TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THERE
;    FILE REFERENCE: 07334-311001
```

RESULT 9
US-10-127-516-4
; Sequence 4, Application US/10127516
; Publication No. US20020187922A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Manji, Gulam A.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-370001
; CURRENT APPLICATION NUMBER: US/10/127,516
; CURRENT FILING DATE: 2002-07-23
; PRIOR APPLICATION NUMBER: US 10/027,629
; PRIOR FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/964,955
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 09/653,901
; PRIOR FILING DATE: 2001-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3857
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (139)...(3240)
US-10-127-516-4

Alignment Scores:
Pred. No.:
Score: 4.66e-81
Percent Similarity: 800.50
Best Local Similarity: 43.64%
Query Match: 28.44%
DB:

Length: 3857
Matches: 219
Conservative: 117
Mismatches: 241
Indels: 193
Gaps: 17

US-10-029-347-2 (1-625) x US-10-127-516-4 (1-3857)

```
Qy  562 -----------ValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeu 576
                    |||||                                    |||||
Db 3073 CTGGCGACCTGGGGGTCATGATGTTCTGTGAAGTTCGAAACAGCAGAGCTGCCTCCTG 3132

Qy  577 GlnThrSerLeuAlaSerValGluLeuSerGluGlnSerLeuGlnGluLeuGlnAla 596
                |||   |||                       :::::
Db 3133 CAGAACCTGGGGTTGCTGCAAATGATATTCAATTATGAGACAAAAAGTGCGTTAGAAACA 3192

Qy  597 ValLysArgAlaLysProAspLeuValIle 606
             :::::          |||||
Db 3193 CTTCAAGAAGAAAGCCTGAGCTGACCGTC 3222
```

RESULT 9

US-10-027-629-4

; Sequence 4, Application US/10027529

; Publication No. US2002197660A1

GENERAL INFORMATION:

; APPLICANT: Bertin, John

; APPLICANT: Manji, Gulam A.

; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN

; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF

; FILE REFERENCE: 07334/341001

; CURRENT APPLICATION NUMBER: US/10/027,629

; CURRENT FILING DATE: 2001-12-22

; PRIOR APPLICATION NUMBER: US 09/964,955

; PRIOR FILING DATE: 2001-09-26

; PRIOR APPLICATION NUMBER: US 09/653,901

; PRIOR FILING DATE: 2000-09-01

; PRIOR APPLICATION NUMBER: US 09/506,067

; PRIOR FILING DATE: 2000-02-17

; NUMBER OF SEQ ID NOS: 17

; SOFTWARE: FastSEQ for Windows Version 4.0

; SEQ ID NO 4

; LENGTH: 3857

; TYPE: DNA

; ORGANISM: Homo sapiens

; FEATURE:

; NAME/KEY: CDS

; LOCATION: (139)...(3240)

US-10-027-629-4

Alignment Scores:

| | |
|---|---|
| Pred. No.: | 1 |
| Score: | 4.66e-81 |
| | 800.50 |
| Percent Similarity: | 43.64% |
| Best Local Similarity: | 28.44% |
| Query Match: | 24.88% |
| Db: | 13 |

| | |
|---|---|
| Length: | 3857 |
| Matches: | 219 |
| Conservative: | 117 |
| Mismatches: | 241 |
| Indels: | 193 |
| Gaps: | 17 |

US-10-029-347-2 (1-625) x US-10-027-629-4 (1-3857)

```
Qy    1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuLeuAspGlyAlaAspGluProAla 20
                    ||||||  ::::  ||||| |:::::  ||||||
Db 1000 ATCGTGAGAACAACCCTCCAGATCCTTCCTCATGGACGGCGTTCGATGAGCTGCAAGGT 1059

Qy   21 LeuGlyGlyProGlnAlaAlaPro---CysThrAspProPheGluAlaAlaSerGlyAla 39
                |||||          :::        |||||  :::::  |||
Db 1060 GCCTTTGACGAGACACATAGGACCGCTGCTGACTGGCCAGGCCAGCGCCGAGCGGGGAGC 1119

Qy   40 ArgValLeuGlyLeuLeuSerLysAlaLeuGluProThrAlaLeuLeuLeuValThr 59
             ::::::         ||||| ||||| |||| |::::::
Db 1120 ATTCCTGAGCAGCCTCATCAGAAGAAAGCCTGAGCTGCTCCCCGAGGCCTCTCTGCTCATCACC 1179

Qy   60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuGlyLeuCysSerProGlnCysAlaGlu 79
             ||||| :::::            ||||| ::::  |||
Db 1180 ACGAGACCTGTGCCGGGACAAACTGGACATGGCAGCACTTCGCTGGACCATCCTGGCCATGTGGAG 1239

Qy   80 ValArgGlyPheSerAspLysAspLysLysTyrPheTyrLysPheTyrArgLysAspAspGlu 99
             :::                |||||            :::::
Db 1240 ATCCTGGGTTCTCCGAGGCCAAAAGGAAAAGGAAGAGTACTTCTTCAAGTACTTCTGTGATGAG 1239

Qy  100 ArgArgGluArgAlaTyrArgGlyPheValLysGlyAsnGlyThrArgGlyPheAlaLeuCys 119
             :::::               |||||           :::::
Db 2299 AGCCACCTCACTTCCAGTTTTCGGGGGCCCTCTTTCAGTTCTGAGCTAGCCACCAGCCAGAGT 2358
```

```
Qy 1300 GCCCAAGCCAGGGCAGCCTCAGTCAGTTCGATTCAGTCGAGGGAACGAGGTCCTTCACCATGTGC 1359

Qy  120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGlnLeuGluLeuGly 139
                  |||||                           :::::
Db 1360 TTCATCCCCTGGTCTGCTGCATGGTTGCACTGGACTGAAACAGCAGCATGGAGAGTGGC 1419

Qy  140 ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuPheHisGlnTyrLeuPheThrSer 159
                :::::       ||||||||:::  :::        ||||||
Db 1420 AAGAGCCTTGCCCCAGACATCAAGACCACCAGCGGGTACGTTCTTCTTCCTTTCCAGT 1479

Qy  160 ValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeu 179
                :::   |||
Db 1480 TTGCTGCAGCCCCGGGAGGGGAGCCCAGAGACACGGCCCTCTGCGCCACCTCTGGGGGCTC 1539

Qy  180 CysArgLeuAlaArgGluGlyValLeuGlyValArgArgAlaGlnAlaGlnAlaLysSerGluLeu 199
                |||||               :::::            ||||||
Db 1540 TGCTCTTTGGCTGCAGATGGAATCTGGAACCAGAAAATCCTGTTGAGGAGTCCGACCTC 1599

Qy  200 GluGlnLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeuSerLysLysGluLeu 219
                 ||||||                        :::      :::::
Db 1600 AGGAATCATGGACTGCAGAAGCGGATGTG---TCTGCTTTCCTGAGGATGAACCTGTTC 1656

Qy  220 ProGlyValLeuGluThrGluValThrTyrGlnValThrGlnSerGlnPheGlnSerProGluPhe 239
                |||||                        ||||||  :::::
Db 1657 CAAAAGGAAGAGTGGACTGCGGAGAGTTCTACAGCTTCATCCACATGACTTCCAGGAGTTC 1716

Qy  240 LeuAlaAlaLeuSerTyrLeuLeuGluAspGlyValProArgThrAlaAlaGlyGly 259
             |||||                        :::            |||||
Db 1717 TTTGCGCCATGTACTACCTGCTGGAGAGAAAGGACGAACGTTCAGGG 1776

Qy  260 ---------------------ValGlyThrLeuLeuArgGlyAspAlaGln-- 269
                                   ||||| ||||| :::::
Db 1777 AGTCGTTGAAGCTCCCAGCCAGACAGCGTGACAGTCCTTCTGAAAACTATGGCAAATTC 1836

Qy  270 ProHisSerHisLeuValLeuThrArgPheLeuPheValProGlyLeuLeuSerAlaGluArg 289
                |||||                        :::::
Db 1837 GAAAAGGGGGTATTTGATTTTGTTGTACGTTTCCTCTTTGGCCTGGTAAACCAGGAGAGG 1896

Qy  290 MetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGlnAla 309
                ||||| :::::                     :::
Db 1897 ACCTCCTACTTGGAGAAGAAATTAAGTTGCAAGATCTCTCAGCAAATCAGGCTGGAGCTG 1956

Qy  310 LeuArgTrpValGlnGlnGlyGlnGlyCysProGlyValValAlaProGluValThrGlu 329
              ::::::             ||||| |||||
Db 1957 CTGGAATGGTTCTACTGTTTGTACGAGATGCAGGAGGAGGACTTCGTGCAA 1960

Qy  330 GlyAlaLysGlyLeuGluAspThrGluLeuGluProGluGluGluGluGluGluGluPro 349
                |||||                        :::
Db 1981 AAAGCTAAAAAAGCTGCAG-----ATCCAGCCCAGCCAG 2013

Qy  350 AsnTyrProLeuGluLeuGluTyrCysLeuTyrGluThrGlnGluValLysAspAlaPheValArg 369
             |||||              :::::
Db 2014 -----------GAATTGTTCTACTGTTTGTACGAGATGCAGGAGGAGGACTTCGTGCAA 2064

Qy  370 GlnAlaLeuGlyAlaPheProGluAlaAlaLeuGlnArgValArgPheCysArgMetAsp 389
             :::::  ||||||          :::            |||||
Db 2065 AGGGCCATGACCATTCCCCAGATTGAGATCAAT-----CTCTCCACCAATGACAC 2118

Qy  390 ValAlaValLeuSerTyrCysValArgCysValSerProAlaGlyLysGlnAlaLeuArgLeuIle 409
             |||||              :::::
Db 2119 CACATGGTTTCTTCCTTTTGCATTGAGACTGTCATCGGGTGGAGTCAGTGTCCCTGGGG 2178

Qy  410 SerCysArgLeuValAlaAlaGlnLeuGlnLeuLysLysSerLeuLysLysArgLeuGln 429
             |||||                        :::::
Db 2179 TTTCTCCATAAACATGCCAAGGAAGGAAGGAGGAGAAAGGAGGCGGACACCTTGAT 2238

Qy  430 ----------AlaSerLeuGlyLeuGlySerGlnGly---ThrThr 441
                      |||
Db 2239 ATGGTGCAGTGTGTCTCCTCCCAAGCTCCTCATGCTGCCTCCGTGTTCCATGGATTGGTGAAC 2298

Qy  442 LysGlnLeuProAlaSerLeuLeuHisProLeuPhe--------------GlnAla 455
             |||||
Db 2299 AGCCACCTCACTTCCAGTTTTCGGGGGCCCTCTTTCAGTTCTGAGCTAGCCACCAGCCAGAGT 2358
```

; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/428,252
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 09/340,620
; PRIOR FILING DATE: 1999-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 4287
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-996-617-5

Alignment Scores:
Pred. No.:                              Length:              4287
Score:                    1.45e-76      Matches:             220
Percent Similarity:       762.00        Conservative:        103
Best Local Similarity:    48.79%        Mismatches:          207
Query Match:              33.23%        Indels:              132
DB:                       33.68%        Gaps:                22
                          10

US-10-029-347-2 (1-625) x US-09-996-617-5 (1-4287)

Qy    1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuLeuAspGlyAlaAspGluLeuPro--  19
Db 1192  ATCCTGTTCAGGCCAGAGAGCGGCTGCTCTCATCCTCGATGGTGAGATGAG---CCAGGA 1248

Qy   20  --AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGly  38
Db 1249  TGGGTCTTGCGAGGAGCCAGTTCTGAGCTCTGTCTGCACTGGAGCCACAGCCGCGGGG 1308

Qy   39  AlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuVal  58
Db 1309  GATGCACTGCTGGGCAGTTTGCTGGGGAAAACTATACTTCCGAGGCATCCTCCTGATC 1368

Qy   59  ThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAla  78
Db 1369  ACGGCTCGGACCCAGCTCTGCAGAACCTCATTCCTCCTTGGAGCAGGCACGTTGGGTA 1428

Qy   79  GluValArgGlyPheSerAspLysAspLeuLysLysTyrPheThrLysPheArgAsp  98
Db 1429  GAGGTCCTGGGGTTCTCAGACAAGGAATATTTCACAGATATTTCACAGAT 1488

Qy   99  GluArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeu 118
Db 1489  GAAAGGCAAGCAATTAGAGCTTTAGGTTGGTCAAACAAACAAAGAGCTCTGGGCCCTG 1548

Qy  119  CysPheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluLeu 138
Db 1549  TGTCTTGTGCCCGGGTGTCCTGGCTGCACTTGCCTGATGCAGCAGATGAAGCGG 1608

Qy  139  GlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThr 158
Db 1609  AAGGAAAAACTCACACTGACTTCCAAGACCACCACACCCTCGTCTACATTACCTTGCC 1668

Qy  159  SerValLeuSerSerAlaProValAlaAspGlyProArgLeuG-nGlyAspLeuArgAsn 178
Db 1669  CAGGCTCTCCAAGCTCCAGCCATTG-----GGACCCCAG-------CTCAGAGAC 1710

Qy  179  LeuCysArgGlyLeuAlaArgGluGlyValLeuGlyArgArgAlaGlnPheAlaGluLysGlu 198
Db 1711  CTCTGCTCTCTGGCTGCTGAGGGCATCTGCAAAAAAGACCCTTTCAGTCCAGATGAC 1770

Qy  199  LeuGluGlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLysGlu 218
Db 1771  CTCAGGAAGCATGGTTAGATGGGGCCATCATCTCCACC---TTCTTGAAGATG------ 1821

Qy  219  LeuProGlyValLeuGluThrGlu--------ValThrTyrGlnPheIleAspGlnSer 235
Db 1822  ------GGTATTCTTCAAGAGCACCCCATCCCTCGAGCTACAGCTTCATTCACCTCTGT 1875

Qy  236  PheGlnGluPheLeuAlaAlaLeuSerTyrLeuLeuGluGluArgArgGlyValProArgThr 255

---

Qy  456  MetThr-----------------------------AspPro------------------  459
Db 2359  CTAACTGAATTGGACCTCAGTGACAATTCTCTGGGGGACCCAGGGATGAGAGTGTTGTGT 2418

Qy  459  -------------------------------------------------------------  459
Db 2419  GAAAACGTCCAGCATCCTGGCTGTAACATTGGAGATGTGTGTTGGGGCGCCTGTGTGGCCTC 2478

Qy  459  -------------------------------------------------------------  459
Db 2479  TCGCATGAGTGCTGCTTCGACATCTCCTTGGTCCTCAGCAGCAACCAGCAGGAGGCTGGTGGAG 2538

Qy  459  -------------------------------------------------------------  459
Db 2539  CTGGACCTGAGTGACACACGCCCTCGGTGACTTCGGAATCAGACTTCTGTGTGTGGGACTG 2598

Qy  460  --------LeuCysHisSerSerLeuThrLeuSerHisCysLysLeuProAlaSerProAspAla  476
Db 2599  AAGCACCTGTTGCGAATCTGAAGAAGCTCTGGTTGGTCAGCTGCTGCCTCACCATCAGCA 2658

Qy  477  ValCysArgAspLeuSerGluGluArgAlaLeuArgAlaProAlaLeuThrGluLeuGlyLeu  496
Db 2659  TGTTGTCAGGGATCTTGCATCAGTATTGAGCACTAGTGAGCCATTCCTGACCAGACTCTATGTG 2718

Qy  497  LeuHisAsnArgLeuSerGlnAlaGlyAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpPro  516
Db 2719  GGGGAGAATGCCTTGGGAGACTTCAGGAGTCGCAATTTTATGTGAAAAAGCCAAGAATCCA 2778

Qy  517  GlnCysArgValGlnThrValArgVal-------------------------------  525
Db 2779  CAGTGTAACCTGCAGAAACTGGGGTGTGGTGAATTCTGGCTCTTACGTCAGTCTGTCTGTTCA 2838

Qy  525  -------------------------------------------------------------  525
Db 2839  GCTTTGCTCCGTACTCAGCACTAATCAGAATCTCACGCACCTTACCTGCGAGGCAAC 2898

Qy  526  GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuVal--GlyMetLeuArgGlnSer  544
Db 2899  ACTCTCCGGAGAC------AAGGGGATCAAACTACTCTGTGAGGGACTTCTGCACCCCGAC 2952

Qy  545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMet---------  561
Db 2953  TGCAAGCTTCAGTGTGTGGAATTAGACAACTGCAACTGCACTCACGTCACACTGCTGCTGGGAT 3012

Qy  561  -------------------------------------------------------------  561
Db 3013  CTTTCCACACTTCTGACCTCCTACCTGCCCAGCCTGCGAAAAGCTGGCGAAAACAATGAC 3072

Qy  562  ----------ValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeu  576
Db 3073  CTGGGGCCACCTGGGGGTTCATGATGTTCTGGAAGTGCTGAAACAGCAGAGCTGCCCCTG 3132

Qy  577  GlnThrLeuSerLeuAlaSerValGluLeuSerGlyGlnSerLeuGlnSerLeuGlnLeuAla  596
Db 3133  CAGAACCTGGGGGTTGTCACAGTATTTCAATATGAGACAGAAAAGTGCGTTAGAAAACA 3192

Qy  597  ValLysArgGlyAlaLysProLeuValAlaIle  606
Db 3193  CTTCAAGAAGAAAGCCTGAGCTGACCGTC 3222

RESULT 10
US-09-996-617-5

; Sequence 5, Application US/09996617
; Patent No. US2002128198A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-340001
; CURRENT APPLICATION NUMBER: US/09/996,617
; CURRENT FILING DATE: 2001-11-27
; PRIOR APPLICATION NUMBER: 09/931,071

```
Qy    565  LeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSerLeuAlaSerVal  584
               |||||        |||      ::||      |||||||||   |||
Db   2824  CTCTGTGAGGGGCTCAGCATCCTGCCTGCCTGCAAACTCATACGCCTGGGGCTGACCAGACA  2883

Qy    585  GluLeuSerGluGlnSerLeuGlnLeuGlnLeuGlnAlaValValAlaArgAlaLysProAspLeu  604
               |||:||      |:: ||        :||:::      |||   |||
Db   2884  ACTCTGAGTGATCGAGATCAGGCAGGCAGGAACTGAGGGCCCTGGAGCAGGAGAAACCTCAGCTG  2943

Qy    605  ValIle  606
               ::|
Db   2944  CTCATC  2949
```

RESULT 11
US-09-931-071-5

; Sequence 5, Application US/0993107l
; Patent No. US20020128219A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; APPLICANT: Alnemri, Emad S.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-335001
; CURRENT APPLICATION NUMBER: US/09/931,071
; CURRENT FILING DATE: 2002-03-18
; PRIOR APPLICATION NUMBER: 09/428,252
; PRIOR FILING DATE: 1999-10-27
; NUMBER OF SEQ ID NOS: 11
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 5
;   LENGTH: 4287
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-931-071-5

Alignment Scores:
```
Pred. No.:                            Length:          4287
Score:                1.45e-76        Matches:          220
Percent Similarity:   48.79%          Conservative:     103
Best Local Similarity: 33.23%         Mismatches:       207
Query Match:          23.68%          Indels:           132
DB:                                   Gaps:              22
```

US-10-029-347-2 (1-625) x US-09-931-071-5 (1-4287)

```
Qy      1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGluAlaAspGluGluPro--  19
               :::|||::|:||::|||||||||||||||||:||||||||||||
Db   1192  ATCCTGTCTAGGCCAGAGCGGCTGCTCTTCATCCTCGATGGTGAGATGAG---CCAGGA  1248

Qy     20  ---AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAla·laSerGly  39
                 |||||||||||||||||||||||||||||||||||
Db   1249  TGGGTCTTGCAGGAGCCCAGTTCAGGTTCGAGCTCTGTCTGCTGCACTGGAGCCCAGCCGCGGCG  1308

Qy     39  AlaArgValLeuGlyLeuGluLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuVal  58
               :::||||||||||||||||||||||
Db   1309  GATGCAGTGCTGGGCAGTTGCTGGGGAAAACTATACTTCCCGAGGCATCCTTCCTGATC  1368

Qy     59  ThrThrArgAlaAlaAlaProGlyLeuGlnGlyArgLeuArgLeuCysSerProGlnCysAla  78
               |||      :||||||||||||||||
Db   1369  ACGGCTCGGACCAGCTCTGCGAGAACCTCATTCCTTCCTTGGACGAGCACGTTGGGTA  1428

Qy     79  GluValArgGlyPheSerAspLysAspLysLysTyrArgPheTyrArgPheArgAsp  98
               |||      :::|||      |||||
Db   1429  GAGGTCCTGGGGTTCTCTGAGTCCAGCAGGAAGAATATTTCACAGAT  1488

Qy     99  GluArgArgAlaGluArgAlaTyrArgPheValGlyAsnGluThrLeuPheAlaLeu  118
               ||||||||||      :::||||||||||
Db   1489  GAAAGGCAAGCAATTAGACCTTTAGGTTGGTCAAATCAAACAAAAGAGCTCTGGGCCCTG  1548

Qy    119  CysPheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGlu... 138
               :::|||      ||||||
Db   1549  TGTCTTGTCCCTGGGTCTCCTGGCTGTCTGGCTGGTTGCCACTTGCCACTGCCGACTG  1608
```

```
Qy   139  GlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThr  158
              :::          :::              ||||||||       :::|||
Db  1609  AAGGAAAAACTCACACTGACTTCCAAGACCACCACCAGCCCTGTCACATTACCTTGCC  1669

Qy   159  SerValLeuSerAlaProValAlaAspGlyProArgLeuGlnGlyLeuAspArgAsn  178
              ||| :::         :::    ||||||           |||    :::
Db  1669  CAGGCTCCCAAGCTCAGCCATTG------GGACCCCAG----------CTCGAGAGAC  1710

Qy   179  LeuCysArgLeuAlaArgGlyLyuValLeuValArgArgAlaAlaGlnAlaAlaLysGlu  198
              |||   ||| :::||| :::          ||||||||
Db  1711  CTCTGCTCTCGCTGCTGAGGGCATCTGCAAAAAAAGAACCCTTTCAGTCCAGATGAC  1770

Qy   199  LeuGluGlnLeuGluGluArgGlySerLysValGlnThrLeuPheLeuSerLysLysGlu  218
              |||                 ||| :::        |||    :::|||
Db  1771  CTCCAGGAGCACTGGGTTAGATGGGGCCATCATCTCCACC---TTCTTGAAGATG----  1821

Qy   219  LeuProGlyValLeuGluThrGlu-----------ValThrTyrGlnPheIleAspGlnSer  235
              ||| :::|||                 :::       ||| :::
Db  1822  ------GGTATTCTTCAAGAGCCACCCCATCCCTCGAGCTACAGCTTCATTCACCTCTGT  1875

Qy   236  PheGlnGluPheLeuAlaAlaLeuSerTyrLeuLeuGluAspGlyAspGlyGlyValProArgThr  255
              ||||||||||||| :::|||  :::       :::
Db  1876  TTCCAAGAGTTCTTTGCAGCAATGTCCTATGCTCTGGAGGA---------------------  1917

Qy   256  AlaAlaGlyGlyValGlyThrLeuLeuArgGlyAspAlaGlnAlaGlnProHisSerHisLeuVal  275
              :::       |||
Db  1918  ---GAGAAGGGGGAGGT-----------------------AAACATTCTAATTGCATC  1950

Qy   276  Leu------------------------------------------------------Thr  277
              |||                                                    :::
Db  1951  ATAGATTTGGAAAAGACCTAGAAGCATATGGAATACATGGCCTGTTTGGGCATCAACC  2010

Qy   278  ThrArgPheLeuPheGlyLeuLeuSerArgAlaGluArgMetArgAspIleGluArgHisPhe  297
              :::|||         ||||||||         |||||||||||||
Db  2011  ACACGTTTCCTATTGGCCTGTTAAGTGATGAGGGGAGAGATGGAGAACATCTTT  2070

Qy   298  GlyCysMetValSerGluArgValLysGlnGlnAlaLeuArgTrpValGlnGlyGlnGly  317
              |||                    :::||||||||
Db  2071  CACTGCGCGGCTGCTTCAG-----GGGAGGAACCTGATGCAGTGGGTC---------  2112

Qy   318  GlnGlyCysProGlyValAlaProGluValThrGluGluLysGlyGlyLeuGluAspThr  337
                                                      ||||||||
Db  2113  ---------------------------CCGTCCCTGCAGCTGCTG  2130

Qy   338  GluGluProGlnGlyGlyGluGlyGlnGlyGluProLeuGluLeuGluTyr  357
              |||       :::         ||||||||||||| :::
Db  2131  CTGCAGCCA--------------CACTCCTGGACCCCTCCAC  2160

Qy   358  CysLeuTyrArgGluThrArgGluGlyAspAlaPheValArgGlnAlaLeuCysSerArgPheProGlu  377
                      :::         :::        ||| :::|||
Db  2161  TGCTTGTACGAGACTCCGGAACAAAACGTTCCTGACACAAGTGATGGCCCATTCGAAGAA  2220

Qy   378  LeuAlaLeuGlnArgValArgPheCys---------------ArgMetAspValAlaValLeuSer  394
              |||                                       :::
Db  2221  ATGGGGCATG--------------TGTGTAGAAACAGAACAGGAGCTTCTTAGTGTGCACT  2265

Qy   395  TyrCysValArgCysCysProCysAlaGlyValAlaLeuArgLeuIleSerCysArg--------  412
              :::          ||| :::|||
Db  2266  TTCTGCATTAAATTCAGCGCCACGTGAAGAGCTTCAGCCGATTGAGGGCAGGCAGCAC  2325

Qy   412  --------------------------------------------------------  412
Db  2326  AGATCAACATGGAAGCCCCACCATGGATCAGTCCTGTTCAGGTGGGTCCCAGTCACAGATGCC  2385

Qy   413  --------LeuValAlaAlaGlnLeuGluGluLysLysLysLysSerLeuGlyTyrLeuArgGlyLeuGln  429
                       :::                   |||
Db  2386  TATTGGCAGATTCTCTTCCGTCCCAGGTCACCAGAAACCTG---AAGGAGCTGGAC  2442

Qy   430  AlaSerLeuGlyGlyGlySerGlnGlyThrThrLysGlnThrLysGlnThrLysSerLeuLeu  449
              |||             :::                          |||
Db  2443  CTAAGT---------GGAAACTCG----------------CTGAGCCACTCTGCAGTG  2475

Qy   450  HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerLeuSerLeuThrLeuSer  469
              |||             :::          :::
Db  1249  TGGGTCTTGCAGGAGCCGGAGTTCTGAGCTCTGTCTGCACTGGAGCCAGCCACAGCCGCCG  1308
```

```
Qy         GlyArgAspLeuSerArgLysGlyThrSerThrSerValTyrLeuPheIleThr  2535
              :::          :::              ||| :::|||
Db  2476  AAGAGTCTTTTGAAGACCCTGAGACGCCCCTGCTGCCTCCTGGAGACCCTGCGGTTGGCT

Qy   470  HisCysLysLysProAspAlaValCysArgAspLeuSerGluArgAlaLeuAlaPro  489
              ||| :::          ||||||||
Db  2536  GGCTGTGGCCTCACAGCTGACTGAGGACTGCAAGGACTTGCTTGGGCTGAGCCAACCAG  2595

Qy   490  AlaLeuThrArgGluLeuGlyLeuHisAsnArgAspLeuSerGluGlyAlaGlyLeuArgMetLeu  509
              ||||||||||| :::         :::        |||
Db  2596  ACCCTGACCAGGGACCTGGACCTGGACCTCAATGTGCTCACGGACGTGAGCCAAACACCTT  2655

Qy   510  SerGlyGlyLeuAlaTrpProGlnProGlnCysArgArgValGlnThrValLeuArgValGlnLeuProAsp  529
              ||| :::                                     :::
Db  2656  TGCCAGAGACTGAGACGCCAGCCTGCAAGCTACAG-----CGACTGCAGCTG--------  2703

Qy   530  ProGlnArgGlyLeu------------------GlnTyrLeuValGlyMetLeuArgGlnSer  544
              |||                                 :::
Db  2704  GTCAGCTGTGGCCTCACGTGCTCCAGCTGTCTCCTGCTCTGCTTAGTGCCAGC  2763

Qy   545  ProAlaLeuThrArgLeuAspLeuSerGlyArgGlnLeuProAlaProMetValThrTyr  564
              |||             :::                  |||
Db  2764  CCCAGCTGAACGAGCTAGACCTGCACCAGCAGCAACACTGGATGACGTTGGCGTGCGACTG  2823

Qy   565  LeuCysAlaValLeuLeuHisGlnGlyTyrGlyGlyLeuGlnThrLeuSerLeuAlaSerVal  584
              |||          :::|||
Db  2824  CTCTGTGAGGGGTCAGGCATCCTGCCTGCAAACTCATACGCCTCGGGGCTGACCAGACA  2883

Qy   585  GluLeuSerGluGluGlnLeuGlnSerLeuGlnLeuGluGlnGlnAlaValLysArgAlaAlaLysProLeu  604
              :::|||
Db  2884  ACTCTGAGTGATGAGTGATGAGGAGGTGAGGAGGAAGCCCTGGAGCAGGAGGAAACCTCAGCTG  2943

Qy   605  ValIle  606

Db  2944  CTCATC  2949
```

```
RESULT 12
US-09-388-221-1
; Sequence 1, Application US/09388221A
; Publication No. US20020192643A1
; GENERAL INFORMATION:
;   APPLICANT: Reed, John C.
;   TITLE OF INVENTION: No. US2002019264 3A1e1 Card Proteins involved in Cell Death Rec
;   FILE REFERENCE: P-LJ 3650
;   CURRENT APPLICATION NUMBER: US/09/388,221A
;   CURRENT FILING DATE: 1999-09-01
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 1
;   LENGTH: 4422
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)..(4422)
US-09-388-221-1

Alignment Scores:
Pred. No.:                          .
Score:                     1.52e-76        Length:        4422
Percent Similarity:        48.79%          Matches:        220
Best Local Similarity:     33.23%          Conservative:   103
Query Match:               23.68%          Mismatches:     207
DB:                        10              Indels:         132
                                           Gaps:            22

US-10-029-347-2 (1-625) x US-09-388-221-1 (1-4422)
```

```
Qy   378  LeuAlaLeuGlnArgValArgPheCys------ArgMetAspValAlaValLeuSer   394
          :::   ||    :::                          ||  :::
Db  2221  ATGGGCATG------TGTGTAGAAACAGACATGGAGCCTTTAGTGTGCACT         2265

Qy   395  TyrCysValArgCysProAlaGlyLeuIleSerCysArg------             412
          ||:||| ::  :|||   ||    :    ||
Db  2266  TTCTGCATTAAATCAGCCGCCACGTGAAGAAGCTTCAGCTGATTGAGGGCAGGCAGCAGAC  2325

Qy   412  ------                                                      412
Db  2326  AGATCAACATGGAGCCCACCATGGTAGTCCTGTTCAGGTGGTGGTCCCAGTCACAGATGCC  2385

Qy   413  ------LeuValAlaAlaGlnGluLysLysLysSerLeuGlyLysLysArgLeuGln   429
                :::   :::        ||||||       :::
Db  2386  TATTGGCAGATTCTTCCGTCCCTCAAGTCACCAGAAACCTG------AAGGAGCTGGAC    2442

Qy   430  AlaSerLeuGlyGlySerGlnGlyThrThrLysProAlaSerLeuLeu           449
          :::                   :::::
Db  2443  CTAAGT------GGAAACTG------CTGAGCCACTCTGCAGTG                 2475

Qy   450  HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSerLeuThrLeuSer  469
          :::      ||                  :::|||     :::
Db  2476  AAGAGTCTTTGTAAGACCCTGAGACGGCCCCTCGCTTCCTGGAGACCCTGCGCTGGCT     2535

Qy   470  HisCysLysLeuProAspAlaValCysArgAspLeuSerGluSerGluAlaLeuArgLeuAlaPro  489
          :::||| :::            |||                :::||||||
Db  2536  GGCTGTGGCCTCACAGCTGAGGACCTTGCCTTGGGCTGAGAGCCAACCAG            2595

Qy   490  AlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeu   509
          :::         ||||||            :::|||       ||||||
Db  2596  ACCCTGACCGAGCTGGACCTGCTGCACAATCGTCCTCGAGCTTCAATGTCTCACGGATGCTGGAGGCCAAACACCTT  2655

Qy   510  SerGluGlyLeuValAlaTrpProGlnCysArgValGlnThrValArgValGlnLeuProAsp  529
          :::   :::            |||              :::            |||
Db  2656  TGCCAGAGACTGAGACAGCCGAGCTGCAAGCTACAG------CGACTGCAGCTG------    2703

Qy   530  ProGlnArgGlyLeu------GlnTyrLeuValGlyMetLeuArgGlnSer           544
          |||                      :::         |||
Db  2704  GTCCAGGTGGGCCTCACGTGGCCTCCAGGACCTGCCGCCTGGCCTGTCCTTAGTGCCAGC    2763

Qy   545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr  564
          |||      :::      |||             :::|||
Db  2764  CCCAGCCTGAAGGAGCTAGACCTGCAGCAGAACAACCTGGATGACGTTGGCGTCCGACTG   2823

Qy   565  LeuCysAlaValValLeuGlnHisGlnGlyLeuGlyLeuAsnThrLeuSerLeuAlaSerVal  584
          |||:::                  :::|||             |||      :::
Db  2824  CTCTGTGAGGGGCTCAGGCATCCTGCCTGCAAACTCATACGCCTGGGGCTGGACCAGACA   2883

Qy   585  GluLeuSerGluGlnSerLeuGlnLeuGluLeuGlnAlaValValArgAlaLysProAspLeu  604
          :::      |||            :::         |||      :::||:::
Db  2884  ACTCTGAGTGATGAGGATGAGAGGGCCCTGGAGCAGGAGGAGAACCTCAGCTG          2943

Qy   605  ValIle                                                       606
          :::|||
Db  2944  CTCATC                                                       2949
```

```
Qy    39  AlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuProThrAlaLeuLeuVal         58
          :::||||||       :::|||||                :::
Db  1309  GATGCACTCTGGGCAGTTGCTGGGAAAACTATACTTCCCGAGGCATCCTTCCTGATC     1368

Qy    59  ThrThrArgAlaAlaAlaProGlyAlaGlyLeuGlnGlyArgLeuCysSerProGlnCysAla  78
          :::|||      :::            |||       :::
Db  1369  ACGGCTCGGACCACCAGCTCTGCAGAACCTCATTCCTTCTTTGGACGAGGCACGTTGGGTA    1428

Qy    79  GlyValArgGlyPheSerAspLysAspLysLysTyrProPheTyrLysPhePheAsp       98
          :::   |||            ||||||       :::
Db  1429  GAGGTCCTGGGGTTCTCTGAGTCCAGCAGGAGGAATATTTCTACAGATATTTCACAGAT     1488

Qy    99  GluArgArgAlaGlnTyrArgPheValLysGluValAsnGluThrLeuPheAlaLeu      118
          :::||||||            :::|||             :::
Db  1489  GAAAGGCAAGCAATTAGAGCCTTTAGGTTGGTCAAATCAAATCAAAGAGCCTTGGCCCCG    1548

Qy   119  CysPheValProPheValCysTrpIleValCysThrValLeuGlnLeuLeuLeu        138
          |||      :::            |||:::
Db  1549  TGTCCTGTGCCCTGGGTGCTGGCTGGCCGACTTGGCCAGATGCAGGCAGATGAAGCGG     1608

Qy   139  GlyArgAspLeuSerArgThrSerThrThrThrSerValTyrLeuLeuPheIleThr     158
          :::||||||      :::|||             :::
Db  1609  AAGGAAAAACTCACACTGACTTCCAAGACACCACAACCCTCTGTCACATTACCTTGCC     1668

Qy   159  SerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyLeuArgAsn    178
          :::|||      :::            ||||||       :::
Db  1669  CAGGCTCTCAAGCTCAGCCATTG------GGACCCAG------CTCAGAGAC          1710

Qy   179  LeuCysArgLeuValLeuAlaArgGlyGlyValLeuGlyArgArgAlaGlnPheAlaGluLysGlu  198
          |||:::            |||::::::            |||
Db  1711  CCCTGCTCTCTGGCTGCTGAGGGCATCTGGGGCATCTGCAAAAAAGACCCTTTCAGTCCAGATGAC  1770

Qy   199  LeuGlnGlnGlnLeuGlnGluLeuArgGlyValGlyThrLeuPheLeuSerLysLysGlu  219
          |||:::            :::|||             |||
Db  1771  CTCAGGAAGCATGGGTAGATGGGCATCTCCACC------TTCTTGAAGATG           1821

Qy   219  LeuProGlyValValLeuGlnThrGlu------ValThrTyrGlnPheIleAspGlnSer  235
          :::|||            :::                  ||||||
Db  1822  ------GGTATTCTTCAAGACACACCCCATCCCTCTGAGCTACCAGCTTCATTCACCCTCTGT  1875

Qy   236  PheGlnGlnLeuPheLeuAlaAlaIleLeuSerTyrLeuLeuGluAspGlyTyrGlyValProArgThr  255
          :::||||||            :::   ||||||             :::
Db  1876  TTCCAAGAGTTCTTTGCAGCAATGTCCTATGTCTTGGAGGAT------GGAGGAAGGGGAGAGGT  1917

Qy   256  AlaAlaArgValGlyValGlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuVal  275
          :::|||      :::                  ||||||       :::
Db  1918  --GAGAGGGGAGAGGT                                              1950

Qy   276  Leu------Thr                                                 277
          :::
Db  1951  ATAGATTTGGAAAAAGACGCTAGAAGCATATGGAATACATGCCCTGTTTGGGGCATCAACC   2010

Qy   278  ThrArgPheLeuPheGlyLeuLeuSerArgAlaGluMetArgAspIleGluArgHisPhe  297
          |||      :::            :::|||             :::
Db  2011  ACACGTTCCATCTGGCCTGTTAAGTGATGAGGGGGAGAGAGGATGGAGAACATCTT       2070

Qy   298  GlyCysMetValSerProGlyValAlaLysValLysGlnGluAlaLeuArgTrpValGlnGlyGln  317
          :::|||            :::            |||             :::
Db  2071  CACTCGCCGGTGTCTCAG------GGGAGGAACCTGATGCCAGTGGGTC------         2112

Qy   318  GlnGlyCysProGlyValAlaProGluValAlaThrGluGluGlyAlaLysGlyLeuGlnAspThr  337
          :::||||||                  :::|||
Db  2113  ------CCGTCCCTGCCAGCTGCTG                                     2130

Qy   338  GluGluProGluGluGluGluGluGluGluProAsnTyrProLeuGluLeuLeuTyr     357
          :::      |||            :::|||
Db  2131  CTGCAGCCA------CACTCTCTGAGTCCCTCCAC                            2160

Qy   358  CysLeuTyrGluThrGlnGluAspLysAlaPheValArgGlnAlaLeuGlnArgPheProGlu  377
          |||:::            ||||||             :::|||
Db  2161  TGCTTGTACGACTCGGAACAAAAACGTTCCTGACACAAGTGATGCCCATTCGAAGAGA     2220
```

```
; NAME/KEY: CDS
; LOCATION: (17)..(4362)
; OTHER INFORMATION: Description of Artificial Sequence: Synthetic
; OTHER INFORMATION: Construct
US-09-388-221-9

Alignment Scores:
Pred. No.:
Score:                 1.59e-76        Length:        4556
                       762.00          Matches:       220
Percent Similarity:    48.79%          Conservative:  103
Best Local Similarity: 33.23%          Mismatches:    207
Query Match:           23.68%          Indels:        132
DB:                                    Gaps:          22

US-10-029-347-2 (1-625) x US-09-388-221-9 (1-4556)
```

Qy      1    MetLeuAlaGlnProGlnLeuPheLeuIleLeuAspGlyAlaAspGluPro---       19
             :::||||
Db    1192   ATCCTGTCTAGGCCAGACGCGGCTGCTCCTTCATCCTCGATGGTGTAGATGAG--CCAGGA   1248

Qy     20    --AlaLeuGlyGlyProGluAlaAlaAlaProCysThrAspProPheGluAlaAlaSerGly    38
               :::||                                    :::
Db    1249   TGGGTCTTGCGAGGACCCAGTTCTGAGCTCTGTCTGCACTGAGCCACCAGCCGGCG    1308

Qy     39    AlaArgValLeuGlyGlyLeuSerLysLysAlaLeuLeuProThrAlaLeuLeuLeuVal    58
             :::||||||    |||    ||||||:||||||    :::|||||||
Db    1309   GATGCCACTGCTGGGCAGTTGCTGGGAAAACTATACTTCCCGAGGCATCCTTCCTGATC   1368

Qy     59    ThrThrArgAlaAlaAlaAlaProGlyArgLeuGlnLeuArgLeuCysSerProGlnCysAla    78
             |||
Db    1369   ACGGCTCGGACCACAGCTCTGCAGAACCTCATTCCTTCTTTGGAGCAGGCACGTTGGGTA    1428

Qy     79    GluValArgGlyPheSerAspLysAspLysLysTyrPheThrLysPhePheArgArgAsp    98
             :::|||    |||
Db    1429   GAGGTCCGTGGGGTTCTCAGCAGGAGGAATAATTTCTACAGATATTTCACAGAT   1486

Qy     99    GluArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeu---   118
             ::::||
Db    1489   GAAAGGCAAGCAATTAGAGCCTTTAGGTGGTCAAATCAAACAAAGAGCTCTGGGCCCTG   1548

Qy    119    CysPheValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGluLeu---   138
             ::::||||||
Db    1549   TGTCTTGTGCCCGGGGTCCTGGCTGGGCCTGCCTGATCGCAGAGAAGCGG   1608

Qy    139    GlyArgAspLeuSerArgThrSerThrThrSerThrValTyrLeuLeuPheIleThr:::   158
             :::||||||
Db    1609   AAGGAAAAACTCACACTGACTTCCAAGACCACCAACCCCTGTCTACATTACCTGCC   1668

Qy    159    SerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnLeuAspGlyArgAsn   178
             ||||||
Db    1669   CAGGCTCCCAAGCTCACGCCATTG-----GGACCCCAG--------CCCAGAGAC   1710

Qy    179    LeuCysArgLeuAlaArgGlyValLeuGlyValLeuGlyArgAlaGlnAlaGlyLysGlu   198
             :::|||
Db    1711   CTCTGCTCCCTGGCTCGCTGGAGGCATCTGCAGAAAAAGACCCCTTTCAGTCCAGATGAC   1770

Qy    199    LeuGluGlnLeuGluLeuArgGlyLeuSerValGlnThrLeuPheLeuSerLysLysGlu   218
             ||||||
Db    1771   CTCAGGAAGCATGGGTTAGATGGGCCATCATCCTCACC---TTCTTGAAGATG------   1821

Qy    219    LeuProGlyValLeuGlnThrGlu-------ValThrTyrGlnPheIleAspAspGlnSer   235
             |||||    :::
Db    1822   ------GGTATTCTTCAAGAGACACCCCATCCCTCTGAGCTACAGCTTCATTACCCTGT   1875

Qy    236    PheGlnGluPheLeuLeuAlaLeuSerTyrLeuLeuGluGluGlyGlyValProArgThr   255
             :::||||||
Db    1876   TTCCAAGAGTCTTTGCAGCAATGTCCATGTCTTGGAGGAT----------------   1917

Qy    256    AlaAlaGlyValGlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuVal   275
             |||
Db    1918   --GAGAAGGGGAGAGGT-----------------AAACATTCTAATTGCATC   1953

Qy    276    Leu---------------------------Thr   277
```

```
Db    1951   ATAGATTTGAAAAGACGCTAGAAGCATATGGAATACATGGCTGTTCGGGGATCAACC   2010
             :::                               |||
Qy    278    ThrArgPheGlyLeuLeuSerAlaGluAlaArgMetArgAspIleGluArgHisPhe   297
             |||||||
Db    2011   ACACGTTCCTATTGGCCTGTTAAGTGATGAGGGGAGAGAGAGATGGAGAAACATCTTT   2070

Qy    298    GlyCysMetValSerGluValLysGlnGluAlaLeuArgTrpValGlnGlyGlyGly   317
             :::|||
Db    2071   CACTCGCGGCGTGTCTCAG-----GGGAGGAACCTGATGCAGTGGGTC   2112

Qy    318    GlnGlyCysProGlyValAlaProGluValThrGlyValAlaLysGlyLeuGluAspThr   337
             |||
Db    2113   ----------CGTCCCTGCAGCTGCTG   2139

Qy    338    GluGluProGluGluGluGluGluGluProAsnTyrProLeuGluGluLeuLeuTyr   357
             :::|||
Db    2131   CTGCAGCCA   2160

Qy    358    CysLeuTyrGluThrGlnGluAspAlaPheValArgGlnAlaLeuAlaLeuCysArgPheProGlu   377
             ||||||
Db    2161   TGCTTGTACAGAGCTCGGAACAAAACGTTCCTGACACAAGTGATGGCCCATTTCGAAGAA   2220

Qy    378    LeuAlaLeuGlnArgValArgPheCys----------ArgMetAspValAlaValAlaValLeuSer   394
             :::||
Db    2221   ATGGGCATG--------TGTTAGAAACATGAGAGCTCTTAGTGTGCACT   2265

Qy    395    TyrCysValArgCysCysProAlaGlyLeuGlnLeuLeuSerCysProGlnCysAla   412
             :::||:||||||||
Db    2266   TTCTGCATTAAATTCAGCCGCCACGTGAAGAAGCTTCAGCTGATTGAGGGCAGGCAGCAC   2325

Qy    412    ------------   412
Db    2326   AGATCAACATGGAGCCCCACCATGGTAGTCCTGTTCAGTGGGTCCCAGTCACAGATGCC   2385

Qy    413    ---LeuValAlaAlaGlnGluGluLysLysLysLysSerLeuGlySerLeuArgLeuGln   429
             :::||
Db    2386   TATTGGCAGATTCTCTTCCGGTCCTCAGGTCACCAGAAACCTG--AAGGAGCTGGAC   2442

Qy    430    AlaSerLeuGlyGlyGlySerSerGlnGlyThrThrLysGlnProAlaSerLeuLeu   449
             :::|||
Db    2443   CTAAGT----------CTGAGCCACTCTGCAGTG   2475

Qy    450    HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSerLeuThrLeuSer   469
             |||
Db    2476   AAGAGTCTTTGTAAGACCCTGGAGACGCCCTCGCTGCCTCCTGAGACCCTGCGGTTGGCT   2535

Qy    470    HisCysLysLysProAspAlaValCysArgAspLeuSerGluLeuArgA:aLeuArgAlaAlaPro   489
             |||
Db    2536   GGCTGGCCTCACGCCTGAGGACTGCAAGGACCTTGCCTTTGGCCTGAGGAGCCAACCAG   2595

Qy    490    AlaLeuThrGluLeuGluLeuHisAsnArgLeuSerGluAlaGluLeuArgGlyLeuArgMetLeu   509
             |||||
Db    2596   ACCCTGACCGAGCGCTGACCTGAGCTTCAATGCTGCTCACGGATGTGGAGCCAAACACCTT   2655

Qy    510    SerGluGlyLeuAlaValArgTrpProGlnThrValArgValGlnLeuProAsp   529
             |||
Db    2656   TGCCAGAGACTGAGACCGAGCTGCAAGCTACAG-------CGACTGCAGCTG------   2703

Qy    530    ProGlnArgGlyGlyLeu------------GlnTyrLeuValGlyMetLeuArgGlnSer   544
             |||||
Db    2734   GTCAACGTTGCGCTCACGCGCTGCCCGCGCTTAGTGCCAGC   2763

Qy    545    ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr   564
             :::||||||
Db    2764   CCCAGCCTGAAGGAGCTAGACCTGCAGCAGAACAACCTGACATTGCCTGCGACTG   2823

Qy    565    LeuCysAlaValLeuGlnHisGlnGlyLeuGlyLeuGlnThrLeuLeuSerLeuAlaSerVal   584
             :::||||||
Db    2824   CTCTGTGAGGGGCTCAGGCATCCTGCCTGCAAACTCATCACGCCTGGGGCTGGACCAGACA   2883

Qy    585    GluLeuSerGluGlnSerLeuGlnLeuGluGlnLeuGlnLeuArgAlaValLysArgAlaLeuProAspLeu   604
             |||||
```

RESULT 14
US-09-996-617-1
; Sequence 1, Application US/09996617
; Patent No. US20020128198A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-340001
; CURRENT APPLICATION NUMBER: US/09/996,617
; CURRENT FILING DATE: 2001-11-2?
; PRIOR APPLICATION NUMBER: 09/932,071
; PRIOR FILING DATE: 2001-08-15
; PRIOR APPLICATION NUMBER: 09/428,252
; PRIOR FILING DATE: 1999-10-27
; PRIOR APPLICATION NUMBER: 09/340,620
; PRIOR FILING DATE: 1999-06-28
; NUMBER OF SEQ ID NOS: 10
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1:
; LENGTH: 5444
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (523)...(4809)
US-09-996-617-1

Alignment Scores:
Pred. No.:
Score:                    762.00
Percent Similarity:       48.79%
Best Local Similarity:    33.23%
Query Match:              23.68%

Length:          5444
Matches:          220
Conservative:     103
Mismatches:       207
Indels:           132
Gaps:              22

US-10-029-347-2 (1-625) x US-09-996-617-1 (1-5444)

```
Qy     1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuPro---  19
            :|:||||||||:||||||||:||||||:||||
Db  1714 ATCCTGTCTAGGCCAGAGGCGGCTGCTCTTCATCCTCGATGGTGAGATGGTGTAGATGA---CCAGGA 1770

Qy    20 --AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGly  38
           ||| |||||||:|
Db  1771 TGGGTCTTCGCAGGAGCCGAGTCTGTCTGCACTGGACCACAGCCGGCG 1830

Qy    39 AlaArgValLeuGlyGlyLeuLeuSerLysAlaAlaLeuLeuProThrAlaLeuLeuVal  58
          ::|||||||||||| ||||||
Db  1831 GATGCACTCGTGGGACTTGCTGGGAAAACTATACTTCCCGGAGCCATCCTTCCTGATC 1890

Qy    59 ThrThrArgAlaAlaAlaAlaProGlyArgLeuGlyArgLeuGlnGlnSerProGlnCysSerAla  78
           ::||||
Db  1891 ACGGCTCGGACCCACAGCTCGCAGAACCTCATTCCTTCTTGGAGCAGGCACGTGGGTA 1950

Qy    79 GluValArgGlyTyrPheSerAspLysAspLysLysTyrPheTyrLysPheArgAspAsp  98
          ||||||||||
Db  1951 GAGGTCTTCGCAGGAGCCGAGTCTGTCTGCACTGGACCACAGCCGGCG 2010

Qy    99 GlyArgArgAlaGluArgAlaThrArgPheValLysTyrArgValAsnGluThrLeuAlaLeu  118
          |||||||
Db  2011 GAAAGGCAAGCAACCTTTAGGTTGTCAAATCAAACAAACAGAGAGCTCTGGGCCCTG 2070

Qy   119 CysPheProValCysTrpIleValCysThrValLeuArgGlnLeuGlnLeuGluLeu  138
          |||||:||
Db  2071 TGTCTTGTGCCCTGGGTGTC-TGGCTGGCCTGCACTTGCCTGATGCAGCAGATGAAGCGG 2130

Qy   139 GlyArgAspLeuSerLysArgThrSerLysThrThrSe-ValTyrLeuLeuPheIleThr  158
          ||||||:||:
Db  2131 AAGGAAAAACTCACACTGACTTCCAGACACCACAACCCTGTCTCACATTACCTGCC 2190

Qy   159 SerValLeuSerSerAlaProValAlaAspGlyTyrProArgLeuLeuGlnGlyLeuAspLeuArgAsn 178
          ::|||
Db  2191 CAGGCTCTCCAAGCTCCAGCCATTG---GGACCCCAG---CTCAGAGAC 2232

Qy   179 LeuCysArgLeuArgArgGluGlyValGlyValLeuArgArgAlaGlnPheAlaGluLysLysGlu  198
          |||||| ||:
Db  2233 CTCTGCTCTCTGGCTGCTGAGGGCATCGGCAAAAAAAGACCCTTTCAGTCCAGATGAC 2292

Qy   199 LeuGluGlnLeuGluLeuArgGlySerLysLysGlnGlnThrLeuPheLeuSerLysLysGlu  218
          ||| ||
Db  2293 CTCAGGAAGCATGGGTTAGATGGGGCCATCATCTCCACC---TTCTTGAAGATG----- 2343

Qy   219 LeuProGlyValLeuGluThrGlu------ValThrTyrGlnPheIleAlaAspPheGlnSer  235
          ||| ||||
Db  2344 ---GGTATTCTTCAAGAGACCACCCCGAGCTACAGCTTCATTCACCTCTGT 2397

Qy   236 PheGlnGluPheLeuAlaAlaLeuSerTyrLeuLeuGluArgProGlyGlyValProArgThr  255
          |||||||:||
Db  2398 TTCCAAGAGTTCTTGGCAGCAATGTCCTATGTCTTGGAGGAT----- 2439

Qy   256 AlaAlaAlaGlyGlyValGlyValGlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuVal  275
          :||||
Db  2440 ---GAGAAGGGGAGAGGT---------AAACATTCTAATTGCATC 2472

Qy   276 Leu-----------------------------------Thr  277
          :::
Db  2473 ATAGATTTGGAAAAGACGCTAGAAGCATATGGAATACATGGCC-GTTGGGCATCAACC 2532

Qy   278 ThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPhe  297
          |||||||||:|
Db  2533 ACACGTTTCCTATTGGGCCTGTTAAGTGATGAGGGGAGAGGAGAGAACACTTT 2592

Qy   298 GlyCysMetValSerGluArgGlnArgValLysGlnGluGlnLeuAlaAlaLeuArgTrpValGlnGlyGly  317
          ||:
Db  2593 CACTGCCGGCGTGTTCTCAG-----GGAGGAACCTGATGCAGTGGGTC 2634

Qy   318 GlnGlyCysProGlyValAlaProGlyValThrGluValThrLeuGlyValAlaAlaLysGlyGluValAspThr  337
          ||:|||
Db  2635 -------CCGTCCCTGCAGTCGCTG 2652

Qy   338 GluGlnProGluGluGluGluGlyGluProAsnTyrProLeuGluLeuGlyLeuTyr  357
          :::|||
Db  2653 CTGCAGCCA-----------CACTCCTCTGGAGTCCCTCCAC 2682

Qy   358 CysLeuTyrGluThrGlnGlnGluAspProAlaPheValArgGlnAlaLeuCysArgProPheProGlu  377
          |||||||
Db  2683 TGCTTGTACGAGACTCAGAACAAAACGTTCCTGACACAAGTGATGCCCATTTCGAAGAA 2742

Qy   378 LeuAlaLeuGlnArgValArgPheCys----------ArgMetAspValAlaValLeuSer  394
          :|: |||
Db  2743 ATGGGCATG------TGTGTAGAACAGACATGGAGCTCTTAGTGTGCACT 2787

Qy   395 TyrCysValArgCysProAlaGlyValGlyGlnAlaGlyLeuLeuSerArgCysArg------  412
          |||
Db  2788 TTCTGCATTAAATTCAGCCGCCACGTGAGGACGCTTCAGCTGATTGAGGGGCAGGCAGGAC 2847

Qy   413 -------LeuValAlaAlaGlnLeuGluGluProGluThrValSerGlyLeuLysLysArgLeuGln  429
          :::::
Db  2848 AGATCAACATGAGCGCCCACCATGGTAGTCCTGTTCAGGTGGGTCCCAGTCACAGATGCC 2907

Qy   430 AlaSerLeuGlyGlySerGlnGlyThrThrLysGlnLeuProAlaSerLeuLeuSer  449
          |||
Db  2908 TATTGGGCAGATTCTTCTTCCGTCCCAAGGTCACCAGAAAACCTG---AAGGAGCTGGAC 2964

Qy   450 HisProLeuProGlnAlaMetThrAspProCysHisLeuSerSerLeuThrLeuSer  469
          ||| :::|||
Db  2965 CTAAGT---------GGAAACTCG-----CTGAGCCACTCTGCAGTG 2997
```

Qy  2884 ACTCTGAGTGATGAGATGAGGGCAGGAACTGAGGGCCCTGGAGCGGAGAGAAAACCTCAGCTG 2943
Qy   605 ValIle 606
Db  2944 CTCATC 2949

Db 20 --AlaLeuGlyGlyProGluAlaAlaAlaProCysThrAspProPheGluAlaAlaSerGly 38

Qy 20 ::: ||| ||| |||
1771 TGGGTCTTGCGAGGAGCCGAGTTCTGAGCTCTGTCTGCCACTGGAGCCAGCCAGCCGGGCG 1830

Db 39 AlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuVal 58

Qy :||||| |||||| |||||| |||||
1831 GATGCACTGCTGGGCAGTTTGCTGGGAAAACTATACTTCCGAGCATCTTCCTGATC 1890

Db 59 ThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAla 78

Qy :||| |||||||
1891 ACGGCCGGACCCAGCTCTGCAGAACCCATTCCTTCTTGGGACAGGCACGTTGGGTA 1950

Db 79 GluValArgGlyPheSerAspLysAspLysLysTyrPheTyrLysPheArgAsp 98

Qy ||||||| |||||||||
1951 GAGGTCGTCTGGGGTTCCAGCCAGCAGGAGGAATATTTCACAGATATTTCACAGAT 2010

Db 99 GluArgArgAlaGluArgAlaTyrArgPheValLysGlnAsnGluThrLeuPheAlaLeu 118

Qy ||||| ::|||:||||||||
2011 GAAAGGCAAGCAATTAGAGCCTTTAGGTGGTCAAATACAAACAAAGAGCCTGGGGCCTG 2070

Db 119 CysPheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluLeu 138

Qy |||||| :|:::||||||
2071 TGTCTTGTGCCCTGCCTGGGTGTCCTGGCTGGCCTGCACTTGCCTGATGCAGAGATGAAGCGG 2130

Db 139 GlyArgAspLeuSerArgThrSerLysThrThrThrSerValTyrLeuLeuPheIleThr 158

Qy ||:|| |||||||||||||||||
2131 AAGGAAAAATCCACACTGGACTTCCAAGACCACCACCACCCCTGTCTACATTACCTTGCC 2190

Db 159 SerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsn 178

Qy |||:|| CTCAGAGAC
2191 CAGGCCTTCCAAGCTCAGCCATTG------GGACCCCAG-------------CTCAGAGAC 2232

Db 179 LeuCysArgArgLeuAlaArgGluGlyValLeuValLeuGlyArgArgAlaGlnPheAlaGluLysGluGlu 198

Qy |||||| ||| ::|||||
2233 CTCTGCTCTCGCTGGCTCTGGAGGGCATCTGGCAAAAAAGACCC-TTCAGTCCAGATGAC 2292

Db 199 LeuGluGlnLeuGluLeuArgGlyValGlySerLysValGlnThrLeuProLeuSerLeuLysLysGlu 218

Qy ||| ::: ||| ||| ||
2293 CTCAGGAAGCATGGTTAGATGGGGCCATCATCTCCACC---TTCTTGAAGATG----- 2343

Db 219 LeuProGlyValValLeuGluThrGlu------- ----ValThrThrTyrGlnPheIleAspPheGlnSer 235

Qy |||||| ::|||||
2344 ------GGTATTCTTCAAGAGACCACCCCATCCCTGAGCTACAGGTTCATTCACCTCTGT 2397

Db 236 PheGlnGlnPheLeuAlaAlaLeuSerTyrLeuLeuGluAspGlyValValValProArgThr 255

Qy ||||||||| :::|||||||||||
2398 TTCCAAGAGTTCTTTGCAGCAATGTCCTATGTCTTGGAGGAT 2439

Db 256 AlaAlaGlyGlyValGlyThrLeuLeuLeuArgGlyAspAlaGlnProHisSerHisLeuVal 275

Qy ||| AACATTCTAATTGCATC
2440 ---GAGAAGGGGGAGGGT----------------AACATTCTAATTGCATC 2472

Db 276 Leu 277
Qy Thr
2473 ATAGATTTGGAAAAGACGCTAGAAGCATATGGAATACATGCCTGTTTGGGCATCAACC 2532

Db 278 ThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPhe 297

Qy ||||||| :|||:||||||||
2533 ACACGTTTCTATTGGGCCTGTTAAGTGATGAGGGGGAGAGAGAATGGAGAAACATCTTT 2592

Db 298 GlyCysMetValSerGluArgValLysGlnGluAlaLeuArgTrpValGlnGlyGlnGly 317

Qy :::|||||| ::::|||||||:|||||
2593 CACTCGCCGGCTGTCTCCAG----GGGAGGAACCTGATGCAGTGGGTC------ 2634

Db 318 GlnGlyCysProGlyValAlaProGluValThrGlnValAlaLysValGlyLeuGluAspThr 337

Qy ||| CCGTCCCTGCCAGGTGCTG
2635 --------------------------------CCGTCCCTGCCAGGTGCTG 2652

Db 338 GluGluProGluGluGluGlyGluGluProAsnTyrProLeuGluLeuLeuTyr 357

Qy |||:|||
2653 CTGCAGCCA---------------------------CACTCTCTGGAGTCCCTCCAC 2682

Qy 358 CysLeuTyrGluThrGlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGlu 377

---

Db 2998 AAGAGTCTTTGTAAGACCCTGAGAGCCCTCGCTGCCTCCTGGAGACCCTCGCGTTGGCT 3057

Qy 470 HisCysLysLeuProAspAlaValCysArgAspAlaLeuGluArgGluAlaLeuArgAlaAlaPro 489

Db ::|||| |||
3058 GGCTGTGGGCCTCACAGCTGAGGACTCAAGGACCTTGCC-TTGGGCTGAGAGCCAACCAG 3117

Qy 490 AlaLeuThrGluLeuGlyLeuLeuHisGluAsnArgGlyLeuSerGluAlaGlyLysGluArgMetLeu 509

Db :||||| |||||||||:|||
3118 ACCCTGACCACGGAGCCTGGACTGAGCTTCAATGTGCCACGGATGCCGGAGCCAAAACCCTT 3177

Qy 510 SerGluGlyLeuAlaTrpProGlnCysArgValGlnThrValArgValGlnLeuProAsp 529

Db ::: |||||||||:||||||||||| CGACTGCAGCTG
3178 TGCCAGAGACTGAGACAGCCGAGTCGCAAGCTACAG-------CGACTGCAGCTG------ 3225

Qy 530 ProGlnArgGlyLeu-----------GlnTyrLeuValGlyMetLeuArgGlnSer 544

Db ||||| :::|||
3226 GTCACGTGTGGCCTCACGTCTGACTCCAGGACCTGGCCTCTGGCTCTGCTTAGTGCCAGC 3285

Qy 545 ProAlaLeuThrThrLeuAspLeuSerArgGlyCysGlnLeuProAlaProMetValThrTyr 564

Db |||:|||||| |||
3286 CCCAGGCCTGAAGGAGCTGAGACCTGCAGCAGACAACCTGGATGGACGTTGGCGACTG 3345

Qy 565 LeuCysAlaAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSerLeuAlaSerVal 584

Db ::::|||||||||||||||
3346 CTCTGTGTGGGGGCTCAGGCATCCTGCCTGCAAACTCATACGCCTGGGGCTGACCAGACA 3405

Qy 585 GluLeuSerGluGlnSerLeuGlnLeuGluLeuGlnAlaValLysArgAlaLysArgProAspLeu 604

Db ||||||| :::::|||:::||||
3406 ACTCAGTGATGAGGATGAGGCAGGAACTGAGGGCCCTGAGCAGCAGGAGAAACCTCAGCTG 3465

Qy 605 ValIle 606

Db :::|||
3466 CTCATC 3471

---

RESULT 15
US-09-931-071-1

Alignment Scores:

| | | | |
|---|---|---|---|
| Pred. No.: | | Length: | 5444 |
| Score: | 2.0e-76 | Matches: | 762.00 |
| Percent Similarity: | 48.79% | Conservative: | 103 |
| Best Local Similarity: | 33.23% | Mismatches: | 207 |
| Query Match: | 23.68% | Indels: | 132 |
| | | Gaps: | 10 |

US-10-029-347-2 (1-625) x US-09-931-071-1 (1-5444)

Qy 1 MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuPro--- 19

Db :::::||||:::::|||||||||||||||||||||||||| |||
1714 ATCCTGCTCTAGGCCAGCAGCGGCTCCTTCATCCTCGATGGTGAGATGAG--CCAGGA 1770

```
Db  2683  ||||||||||||||||||||    :::    |||    |||
          TGCTTGTACGAGACTCGGAACAAAACGTTCCTGACACAAGTGATGGCCCATTCGAAGAA  2742

Qy   378  LeuAlaLeuGlnArgValArgPheCys--------ArgMetAspValAlaValLeuSer  394
          :::     |||    :::    |||:::      |||
Db  2743  ATGGGCATG--------TGTGTAGAAACAGACATGAGCTTAGTGTGCACT  2787

Qy   395  TyrCysValArgCysCysProAlaGlyGlnAlaLeuArgLeuSerCysArg------  412
          :::::|:|:   :::    |||:::|||||
Db  2788  TTCTGCATTAAATTCAGCGCCGCCACGTGAGAAGCTTCAGCTGATTGAGGGCAGGCAGGAC  2847

Qy   412  ------  412
Db  2848  AGATCAACATGGAGCCCACCATGTAGTCCTGTTCAGGTGGTCCCAGTCACAGATGCC  2907

Qy   413  ------LeuValAlaAlaGlnGlu:LysLysLysSerLeuGlyLysGlyArgLeu:Gln  429
          :::::   :::    ||:    |||
Db  2908  TATTGGCAGATTCTTCTTCCGTCCCAAGGTCACCAGAAACCTG--AAGGAGCTGGAC  2964

Qy   430  AlaSerLeuGlyGlySerGlnGlyThrThrLysGlnLeuProAlaSerLeuLeu  449
          |||   :::    |||
Db  2965  CTAAGT-------GGAAACTCG--------CTGAGCCACTCTGCAGTG  2997

Qy   450  HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSerLeuThrLeuSer  469
          :|:|    |||    |||:::
Db  2998  AAGAGTCTTTGTAAGACCCTGGAGACGCCCTCGTGCCTCCTGGAGACCCTGCGGTTGGCT  3057

Qy   470  HisCysLysLysLeuProAspAlaValCysArgAspLeuSerGluAlaLeuArgAlaAlaPro  489
          |||:::|||    :::|||||||
Db  3058  GGCTGTGGCCTCACAGCTGAGGACCTTGCCTTTGGGCTGGAGAGCCAACCAG  3117

Qy   490  AlaLeuThrGluLeuGly:LeuLeuHisAsnArgLeuSerGlnAlaGlyLeuAlaGlyMetLeu:  509
          |||    :::    ||||
Db  3118  ACCCTGACCGAGCTGACCGAGCTTCAATGTGCCACGGATGCTGGAGCCAAACACCTT  3177

Qy   510  SerGluGlyLeuAlaTrpProGlnCysArgValGlnThrValArgValGlnLeuProAsp  529
          |||    :::    |||:::|||
Db  3178  TGCCAGAGACTGAGACAGCCGAGCTGCAAGCTACAG------CGACTGCAGCTG------  3225

Qy   530  ProGlnArgGlyLeu--------GlnTyrLeuValGlyMetLeuArgGlnSer  544
          ||||||    :::   |||
Db  3226  GTCAGCTGTGGCCTCACGGCCTCACGTCGCTCGACCTGCCCTCGTGCTTAGTGCCCAGC  3285

Qy   545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr  564
          ||:::|||    :::|||
Db  3286  CCCAGCCTGAAGAGGTGACGTAGACTGCGAGCAGAACAACCTGGATGACGTTGCGTGCGACTG  3345

Qy   565  LeuCysAlaValLeuGluGlnHisHisGlnGlyCysGlyGlyLeuGlnThrLeuSerLeuAlaSerVal  584
          |:|||    |||:::    |||
Db  3346  CTCTGTGAGGGGCTCAGGCATCCTGCCTGCAAACTCATACGCCTGGGGCTGGACCAGACA  3405

Qy   585  GluLeuSerGluGlnSerLeuGlnGluLeuGlnArgAlaValLysArgAlaLysProAspLeu  604
          ||||||    :::    |||
Db  3406  ACTCTGAGTGATGAGATGAGAGCAGGAACTGAGGGGCCCTGGAGCAGGAGGAAACCTCAGCTG  3465

Qy   605  ValIle  606
          :::|||
Db  3466  CTCATC  3471
```

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

CM nucleic - nucleic search, using sw model

Run on: October 3, 2003, 16:57:40 ; Search time 644 Seconds
(without alignments)
10651.043 Million cell updates/sec

Title: US-10-029-347-1
Perfect score: 2689
Sequence: cggacgcgtgggcgcgagc........aaaaaaaaaaaaaaaaa 2689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1708419 seqs, 1275431651 residues

Total number of hits satisfying chosen parameters: 3416838

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications NA:*
1: /cgn2_6/ptodata/1/pubpna/US07_PUBCOMB.seq:*
2: /cgn2_6/ptodata/1/pubpna/PC7_NEW_PUB.seq:*
3: /cgn2_6/ptodata/1/pubpna/US06_NEW_PUB.seq:*
4: /cgn2_6/ptodata/1/pubpna/US06_PUBCOMB.seq:*
5: /cgn2_6/ptodata/1/pubpna/US07_NEW_PUB.seq:*
6: /cgn2_6/ptodata/1/pubpra/PCTUS_PUBCOMB.seq:*
7: /cgn2_6/ptodata/1/pubpna/US08_NEW_PUB.seq:*
8: /cgn2_6/ptodata/1/pubpna/US08_PUBCOMB.seq:*
9: /cgn2_6/ptodata/1/pubpna/US09A_PUBCOMB.seq:*
10: /cgn2_6/ptodata/1/pubpna/US09B_PUBCOMB.seq:*
11: /cgn2_6/ptodata/1/pubpna/US09C_PUBCOMB.seq:*
12: /cgn2_6/ptodata/1/pubpna/US09_NEW_PUB.seq:*
13: /cgn2_6/ptodata/1/pubpna/US10A_PUBCOMB.seq:*
14: /cgn2_6/ptodata/1/pubpna/US10B_PUBCOMB.seq:*
15: /cgn2_6/ptodata/1/pubpna/US10_NEW_PUB.seq:*
16: /cgn2_6/ptodata/1/pubpna/US60_NEW_PUB.seq:*
17: /cgn2_6/ptodata/1/pubpna/US60_PUBCOMB.seq:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | II | | Description |
|---|---|---|---|---|---|---|---|
| 1 | 1708.8 | 63.5 | 2844 | 11 | US-09-965-621-17 | | Sequence 17, Appl |
| 2 | 277 | 10.3 | 3186 | 14 | US-10-066-521-17 | | Sequence 17, Appl |
| 3 | 273.4 | 10.2 | 3109 | 11 | US-09-965-621-23 | | Sequence 23, Appl |
| 4 | 228 | 8.5 | 2524 | 11 | US-09-965-621-58 | | Sequence 58, Appl |
| 5 | 226.4 | 8.4 | 3102 | 12 | US-10-132-967-6 | | Sequence 6, Appli |
| 6 | 226.4 | 8.4 | 3102 | 13 | US-10-127-516-6 | | Sequence 6, Appli |
| 7 | 226.4 | 8.4 | 3102 | 13 | US-10-027-629-6 | | Sequence 6, Appli |
| 8 | 226.4 | 8.4 | 3857 | 12 | US-10-132-967-4 | | Sequence 4, Appli |
| 9 | 226.4 | 8.4 | 3857 | 13 | US-10-127-516-4 | | Sequence 4, Appli |
| 10 | 226.4 | 8.4 | 3857 | 13 | US-10-027-629-4 | | Sequence 4, Appli |
| 11 | 200.4 | 7.5 | 1650 | 10 | US-09-833-381-1976 | | Sequence 1976, Ap |
| 12 | 179.2 | 6.7 | 3099 | 12 | US-10-132-967-3 | | Sequence 3, Appli |
| 13 | 179.2 | 6.7 | 3099 | 13 | US-10-127-516-3 | | Sequence 3, Appli |
| 14 | 179.2 | 6.7 | 3099 | 13 | US-10-027-629-3 | | Sequence 3, Appli |
| 15 | 179.2 | 6.7 | 3431 | 12 | US-10-132-967-1 | | Sequence 1, Appli |
| 16 | 179.2 | 6.7 | 3431 | 13 | US-10-127-516-1 | | Sequence 1, Appli |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| 17 | 179.2 | 6.7 | 3431 | 13 | US-10-027-629-1 | | Sequence 1, Appli |
| 18 | 179 | 6.7 | 4931 | 12 | US-10-028-374-1 | | Sequence 1, Appli |
| 19 | 179 | 6.7 | 4931 | 12 | US-10-183-770-1 | | Sequence 1, Appli |
| 20 | 177.6 | 6.6 | 3189 | 12 | US-10-239-663-10 | | Sequence 10, Appl |
| 21 | 171.6 | 6.4 | 2940 | 9 | US-09-848-035-14 | | Sequence 14, Appl |
| 22 | 171.6 | 6.4 | 2940 | 10 | US-09-986-224-14 | | Sequence 14, Appl |
| 23 | 171.6 | 6.4 | 3150 | 12 | US-10-239-663-9 | | Sequence 9, Appli |
| 24 | 171.6 | 6.4 | 3263 | 9 | US-09-848-035-12 | | Sequence 12, Appl |
| 25 | 171.6 | 6.4 | 3263 | 10 | US-09-986-224-12 | | Sequence 12, Appl |
| 26 | 167 | 6.2 | 2982 | 14 | US-10-066-521-25 | | Sequence 25, App. |
| 27 | 167 | 6.2 | 2985 | 11 | US-09-965-621-15 | | Sequence 15, Appl |
| 28 | 167 | 6.2 | 3368 | 14 | US-10-066-521-23 | | Sequence 23, Appl |
| 29 | 165.4 | 6.2 | 2575 | 9 | US-09-848-035-7 | | Sequence 7, Appli |
| 30 | 165.4 | 6.0 | 497 | 13 | US-10-027-632-133904 | | Sequence 133904, |
| 31 | 160.6 | 6.0 | 1515 | 13 | US-10-027-632-133904 | | Sequence 133904, |
| 32 | 141.8 | 5.3 | 1515 | 10 | US-09-833-381-1904 | | Sequence 1904, Ap |
| 33 | 141.8 | 5.3 | 1515 | 10 | US-09-833-381-1905 | | Sequence 1905, Ap |
| 34 | 139.4 | 5.2 | 5100 | 11 | US-09-956-712-12 | | Sequence 12, Appl |
| 35 | 137.8 | 5.1 | 4200 | 10 | US-09-388-221-3 | | Sequence 3, Appli |
| 36 | 137.8 | 5.1 | 4287 | 10 | US-09-996-617-5 | | Sequence 5, Appli |
| 37 | 137.8 | 5.1 | 4287 | 10 | US-09-931-071-5 | | Sequence 5, Appli |
| 38 | 137.8 | 5.1 | 4332 | 10 | US-09-388-221-5 | | Sequence 5, Appli |
| 39 | 137.8 | 5.1 | 4422 | 10 | US-09-388-221-1 | | Sequence 1, Appli |
| 40 | 137.8 | 5.1 | 4466 | 10 | US-09-388-221-11 | | Sequence 11, Appl |
| 41 | 137.8 | 5.1 | 4556 | 10 | US-09-388-221-9 | | Sequence 9, Appli |
| 42 | 137.8 | 5.1 | 5444 | 10 | US-09-996-617-1 | | Sequence 1, Appli |
| 43 | 137.8 | 5.1 | 5444 | 10 | US-09-931-071-1 | | Sequence 1, Appli |
| 44 | 137.8 | 5.1 | 5444 | 11 | US-09-956-712-3 | | Sequence 3, Appli |
| 45 | 137.8 | 5.1 | 6531 | 11 | US-09-956-712-11 | | Sequence 11, Appl |

ALIGNMENTS

RESULT 1
US-09-965-621-17
; Sequence 17, Application US/09965621
; Publication No. US20030077699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 17
; LENGTH: 2844
; TYPE: DNA
; ORGANISM: Homo sapien
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (1)...(2841)
US-09-965-621-17

Query Match 63.5%; Score 1708.8; DB 11; Length 2844;
Best Local Similarity 90.3%; Pred. No. 0;
Matches 1309; Conservative 0; Mismatches 37; Indels 168; Gaps 1;

Qy 7 CGTGGGCGGCGAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGGCGGCCGCCGG 66
       |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 731 CGGGCACGCGCGAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCAGTGCCCCGG 790

Qy 67 TGCCGCAGATGCTGGCCCAGCCGCAGCGGCGCTGCTCTTCATCCTGGACGGGCCGGACGAGC 126

This page contains a nucleotide sequence alignment (Query "Qy" vs. Database "Db") with residue position numbers. The alignment is presented in the dense sequence format typical of patent sequence listings.

APPLICANT: Wang, Weiye
APPLICANT: Blatcher, Maria
TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN/NBS/LRR
TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
FILE REFERENCE: 07334-334001
CURRENT APPLICATION NUMBER: US/10/066,521
CURRENT FILING DATE: 2002-06-25
PRIOR APPLICATION NUMBER: 60/318,645
PRIOR FILING DATE: 2001-09-10
PRIOR APPLICATION NUMBER: 60/265,231
PRIOR FILING DATE: 2001-01-31
NUMBER OF SEQ ID NOS: 25
SOFTWARE: FastSEQ for Windows Version 4.0
SEQ ID NO 17
LENGTH: 3186
TYPE: DNA
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: CDS
LOCATION: (1)...(3183)
US-10-066-521-17

Query Match 10.3%; Score 277; DB 14; Length 3186;
Best Local Similarity 48.9%; Fred. No. 2.1e-58;
Matches 891; Conservative 0; Mismatches 915; Indels 15; Gaps 5;

Qy 6 GGGTGGGGCGCCAGCCTGCTGACCTGATCCTGACCAGTGCCCGACCGCCGCGGCCG 65
Db 775 GCCACGGAATGCAGGCATGCAAGACCTCATCTTCAGCTGCTGGCCTGAGCCCAGCGCGCCT 834

Qy 66 GTGCCCAGATGCTGGCCCCAGCCCGCCCCCGAGGCGGCTGGCGTCTCCATCCTGGACGGAGACGGCGGGACGAG 125
Db 835 CTCCAGGAGCTCATCCGAGTCCCGAGTGGGCCCTCCTTTTCATCATCGACGGCTTCGATGAG 894

Qy 126 CT--GCCGGCGGTGGGGGGTCCCGAGCCCGGCCCTGCACAGACCCCTTGCAGGCGGCG 182
Db 895 CTCAAGCCCTTTCTTCCACGATCCTCAGGGACCCTGGTGCCTCTGCGGGAGGAAAACGG 954

Qy 183 AGCCGCGCGGGGTGCTAGGCGGTGCTGCTGAGTAGGCGGCTGCTGCCCACGGGCCCTCCTG 242
Db 2029 CGCGCGAGGTGCTCCCGCAGGAGCGCACACGCTGTTGGTGCAGCTACCAGGAGGACCGTT 2088

Qy 955 CCCACGGAGCTGCTTCTTAACAGCTTAATTCGGAAGAAGCTGCTCCCTGAGCTATCTTTG 1014
Db 955 ...

Qy 243 CTGGTGACCACGCGCGGCCCGCGCCGCCCGGAGGCTGCAGGGCCGCTGTGTTCCCGCAG 302
Db 2089 CTGTTGGACGCCTACAGTGAACATCTGGCAGCGGCCCTGCGCCACCAATCCAAACCTGATA 2148

Qy 1015 CTCATCACCACCACGCCCACACGGCCCCACCGGTCTGGGCTTGGAAGCTCCACCGTCGTCCTGCTGGAGCACCCCCAGG 1074
Db 1380 AGTTCTCAAGGCACCACCACAAAACAACTGCCAGCCCTTCCCTTCATCCCACTCTTTCAGGCA 1439

Qy 303 TGGCCCGAGCGCGCGGTGCCAGGCGCGCCTACCGGGCCTACCGCGTTCGTGAGGAGAACGAGATTTCTACAAGTTCTTC 362
Db 2149 GAGCTGTCTCTGTACCGAAATGCCCTGGGCAGCCGTGAGCGCTGCTCTGTCAAGGA 2208

Qy 1075 CATGTGGAGATCCTGGGCTTCTGGGCTGGGCTGGGGGCAAGTCTTCAATTACGTGAGGGACAACGAGCCTCGTATTTC 1134
Db 1440 ATGACTGACCCACTGTCCATCGAGCAGGCTTCAGAGGCTTCAGAACCTGGAGCTTCCATAGCCAATAAGAATTTGACAAGGATGGAT 1499

Qy 363 CGGGATGAGGAGGCGGCCCAGCGCGCGCCTACCGGGCCCTACCGGCGTCGTTGAGGAGAACGAGACGGCGGTTC 422
Db 2209 CTCAGACACCCCAACTTCAGAACTTCAGAACCTGGAGCTTCAGAGGTGCGCATCTCCAGC 2268

Qy 1135 CACAATGCAGAGCAGGCGGGGCCAAGTCTTCAATTACGTGAGGGACAACGAGCCTCGTCTTC 1194
Db 1500 GCGGTGCTGCCGGAGAACCTTTCTGGTGCTGGATCGTGCTGGATCGCACCGGTGCGGCCCAGCAGCTG 1559

Qy 423 GCGCTGTGTCTGTTCGTGCCCTTCGTGTGCTGGATCGGTGCATCGTGGTGCCACGGTGCGGCCCAGCAGCTG 482
Db 2269 TCAGCCTGCGGAGGACCTCGTCTCTGCAGCTGCTGTGTACCTGGTGTGTTGACCAATAGCTGAGTGGAGGGCCTAGCCTGG 2328

Qy 1195 ACCATGTCGTCGTCCCCGTGGTGTGCTGGGTGCTGTGTACCTGGTGTGGCTGTACCTGCTCCAGCAGCAGCTG 1254
Db 1560 CTCTCCCACAACAGGCTCAGTGAGCGGGACTCAGTGACTGAGTGAGGGCTAGCCTGG 1619

Qy 483 GAGTCGCGGTCGGACCTGTCGGCGCCACGTCGGTCGTGGAGACGAACCTGTCAGTGTACCTGCTTTTC 542
Db 2329 CTCAGTGGCACGCGGCCTTGGCGCATGCGTGCTGTTTCGGAGGGGGCCTGCGGGCAT 2388

Qy 1255 GAGGGGTTGTGCTCTTGCTGGGCGCAGATGGCGTCTGGGAATCAGAAAATCCTATTTGAGGAG 1314
Db 1620 CGGGCAGTGGCAGGGTGCAGACGGTCAGGGTACAGCTGCTGCTCGACCCCCAGCCAGGGGCTC-- 1676

Qy 543 ATCACCAGGTTCTGAGCTCG3CTCCGGTAGCCCGGTAGCCGGCCCGGTTCAGGCGGACCTG 602
Db 2389 CCCCAATGCAGGCTGCCAGATGATTCAGTTGAGGAAGTGTCACCCTGGAGTTCGGGGCTTGT 2448

Qy 1315 CTGCTGAGTCTGATGCAACCCAAGCCCACCCCAAGCCGGGGGCCCTCCAGGCGGGGGCCCTCCAGGCACCCCAACCAG 1374
Db 1677 CAGTACCTTGGTGGGTATGCTTCGGCAGAGCCCCGCAGTTTTGCCGAG 1736

Qy 603 CGGCAATCTGTGCCGCCTGGCCGCGAGGGCGTCCTCGGACGGCAGGGGCAGTGCGGAG 662
Db 2449 CAGGAGATGGCTTCGTGTCTTGGCACCACCACATCTGTTGAGTTGACGTTGACCTGACAGGA 2508

Qy 1375 AGAGGGTTGTCTCTTGCGG3CAGATGGCGTCTGGAATCAGAAAATCCTATTTGAGGAG 1434
Db 1737 TGCCAACTGCCCGCCCCCATGGTGACCTACCTGTGTCGAGTCTTGCAGCACCAGGAT 1796

Qy 663 AAGGAACTGAGCAACTGGAGCTTCGTGGCTCAAAGTGCAGACGCTGTTTCTCAGCAAA 722
Db 1435 CAGGACCTCCGGAAGCAGCACGGCCT--AGACGGGGAAGACGTCTCTGCCTTCCTTCCTCAACATG 1491

Qy 1435 [continued]
Db 2509 AATGCACTGAGGAGATTTGGGCCTGAGGTTACTATGCCAGGGACTGAGGCACCCCAGTCTGC 2568

Qy 723 AAGGAGCTGCCGGGCGTGCTGGGAGACAGAGGTCACTCACCTACCAGTTCATTGACCAGAGCTTC 782
Db 1492 AACATCTTCAGACAGACATCAACTGTGAGAGGTACTACAGCTTCATCCACTTGAGTTC 1551

Qy 783 CAGGAGTTCCTCGCCGCACTGTCCTACCTGCTGAGGAGGACGGCCGGGTGCTGCCAGGACCGGC 842
Db 1552 CAGGAATTCTTGCAGCCTATATCCTGGACGAGGAGGGCGGGGGCAGGCCCA 1611

Qy 843 GCTGGCGACGGTTGGACACCTCCTGCTGTGGGACGGCCAGCCGCA---CAGCCACTTGGTG 899
Db 1612 GACCAGAGGTGACCAGGCTGTTGACCGAGTACGCGTTTCTGAAAGAGCTTCCTGGCA 1671

Qy 900 CTCACCACGCGCGTTCCTCTTCGGACTGCTGGAGCCGCTGGAGGCGGGATGCGGAGCCCGAGGGC 959
Db 1672 CTCACCACGCCGCTTCCTGTTTGGACTCCTGAACAGGAGAGACAGGAGCCCACCTGGAGAAG 1731

Qy 960 CACTTCGGCTCATGGTTCAGAGCGTGAAGCAGGACGCTGTGAAGCAGAGGCCCTGCGGTGGGTCAGGGA 1019
Db 1732 AGTCTCTGTGGAAGGTCTCGCCGCACATCAAGATGACCCTGTTGCAGTGGATCCAAAGC 1791

Qy 1020 CAGGGACAGGGCTGCCCCGGAGTGCACCCAGAGTGACCGAGGTCGAG 1079
Db 1792 AAAGCTCAGAGCCAGACGGCTCCACCCGCAGCAGGGCTCCTTGGAGTTCTTCAGCTGCTTG 1851

Qy 1080 GACACCGACGAGCAGCAGGAGGAGGAGGACGAGGAGCCCAACTACCCAACTACCCACTCGAGTTG 1139
Db 1852 TACGAGATCCAGGAGACGAGGAGGAGGTTTATCCAGCAGGGTTTGCGCCGACGGCCTGATCGTG 1911

Qy 1140 CTGTACTGCTGTACGAGACGCAGGAGGACAGGAGCACGCGTTGTGCGCCAAGCCCTGTGCCGGTTC 1199
Db 1912 GTCAGCAACATTGCCTCCAAGATGGAGCACATGTCTCCTCGTGTCTGTCGCTCGAAGCGCTGC 1971

Qy 1200 CCGGAGCTGGCGTCTGCAGCGAGTGGCTTCTGCTGCGGCCCTGCAGCACCCCTCGAGGCGGCGCG 1259
Db 1972 AGGAGGCCCCAGGTGCTGCACTTG---TATGGCGGCCACCTACAGCGCGGGGGAAGAC 2028

Qy 1260 TGCGTGAGGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCT 1319
Db 1319 ...

Qy 1320 GCGCGAGAAGAGAGAGAGAGAGCCTGGGAGGCGCTGGGAGGCCCTGCTGCTGGTGGCGGC 1379
Db 2089 ...

```
Qy    1797 GGCCTGCGAGACCCTCAGTCTG 1817
           ||||||| ||| ||| |||||||
Db    2569 AGACTACGGACTTTGTTGGCTG 2589

RESULT 3
US-09-965-621-23
; Sequence 23, Application US/09965621
; Publication No. US2003007769SA1
; GENERAL INFORMATION:
;  APPLICANT: Reed, John C.
;  APPLICANT: Godzik, Adam
;  APPLICANT: Chu, Zhi-Liang
;  APPLICANT: Pawlowski, Krzysztof
;  APPLICANT: Fiorentino, Loredana
;  APPLICANT: Ariza, Maria Eugenia
;  APPLICANT: Stehlik, Christian
;  TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
;  TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
;  FILE REFERENCE: P-LJ 4816
;  CURRENT APPLICATION NUMBER: US/09/965,621
;  CURRENT FILING DATE: 2001-09-25
;  PRIOR APPLICATION NUMBER: US 09/671,760
;  PRIOR FILING DATE: 2000-09-26
;  NUMBER OF SEQ ID NOS: 64
;  SOFTWARE: FastSEQ for Windows Version: 4.0
;  SEQ ID NO 23
;  LENGTH: 3108
;  TYPE: DNA
;  ORGANISM: Homo sapien
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)..(3105)
US-09-965-621-23

Query Match        10.2%;   Score 273.4;  DB 11;  Length 3108;
Best Local Similarity  49.5%;  Pred. No. 1.6e-57;
Matches  304;  Conservative   0;  Mismatches 301;  Indels  20;  Gaps  7;
```

(DNA sequence alignment blocks, Qy and Db rows, positions 6–2306)

```
Qy   1616 CTGGCCGCAGTGCAGGGTGCAGACGGTCAGGTACAGCTGCCTGACCCCAGCGCGAGGGCT 1675
          ||||||| |||| |  ||||||   |  ||||||||||| | || | ||||| ||  |||
Db   2307 GCATCCCCAGTGCAGGGCTGCAGCGATGATTCAGTTGAGGAAGTGTCAGTGGAGTCCGGGGC 2366

Qy   1676 C---CAGTACCTCGTGGGCATGCGTTCGGCGACCAACCCACCATCTGGTCACCCTGGATCTCAG 1732
          |    |  || |    |||||| | ||  |||||||||| || || | |  ||   |||
Db   2367 TTGTCAGGAGATGCTTCTGCGTCGGCGACCAACCCACATCTGGTTGAGTTGGACTGAC 2426

Qy   1733 CGGCTGCCAACTGCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCGCAGCACCAGGG 1792
          | |||||     ||| || ||||   |  || || |
Db   2427 AGGAAATGCACTGAGGATTGGGCCTGAGGTTACTATGCCAGGAGGCACCCCAGT 2486

Qy   1793 ATGCGGCCTGCAGACCCTCAGTCTG 1817
          ||  ||||||| |||||| |
Db   2487 CTGCAGACTACGGACTTTCTGGCTG 2511
```

RESULT 4
US-09-965-621-58
; Sequence 58, Application US/09965621
; Publication No. US20030077699A1
; GENERAL INFORMATION:
; APPLICANT: Reed, John C.
; APPLICANT: Godzik, Adam
; APPLICANT: Chu, Zhi-Liang
; APPLICANT: Pawlowski, Krzysztof
; APPLICANT: Fiorentino, Loredana
; APPLICANT: Ariza, Maria Eugenia
; APPLICANT: Stehlik, Christian
; TITLE OF INVENTION: PAAD Domain-Containing Polypeptides,
; TITLE OF INVENTION: Encoding Nucleic Acids, and Methods of Use
; FILE REFERENCE: P-LJ 4816
; CURRENT APPLICATION NUMBER: US/09/965,621
; CURRENT FILING DATE: 2001-09-25
; PRIOR APPLICATION NUMBER: US 09/671,760
; PRIOR FILING DATE: 2000-09-26
; NUMBER OF SEQ ID NOS: 64
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 58
; LENGTH: 2524
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (176)...(2332)
; NAME/KEY: misc_feature
; LOCATION: (306)...(458)
; OTHER INFORMATION: n = A,T,C or G
US-09-965-621-58

Query Match     8.5%;    Score 228;    DB 11;    Length 2524;
Best Local Similarity  56.4%;   Pred. No. 2.5e-46;
Matches  467;   Conservative  0;   Mismatches 355;   Indels 6;   Gaps 2;

```
Qy     14 GCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCGACCGCGGCCGCCGTGCCGCA 73
          |||||||||||| |||| |||| |||| ||||||||| ||| |||||||||||||
Db    982 GAGGAGCCTGGGGGACCGCAGCGCGCGATCATGAGCTGCGCGATCATGAGCTGCGC 1041

Qy     74 GATGCTGGCCGAGAAAACCCTCCAGAATCCTCTTCCCTATGGACCGGCTTCGATGAGCTGCAAGG 130
          |||||||||| ||||||||| |||||||||||||||||||||||||| ||| | |
Db   1042 GATGCTGGCCGAGAAAACCCTCCAGAATCCTCTTCCCTATGGACGGGCGGTTGCAGGGTTCGATGAGCGGCGCAAGG 1101

Qy    131 GGGCGCTGGGGGGCCCCCGAGGCGCCCTTCGAGGCGGCGGAGCGGCGCC 190
          ||||| | ||| | || |||||| |||||| |||||| || | |
Db   1102 TGCCTTTGACGAGCACATAGGACCGCTCGACTGACTGGCGAGAAGGCCGAGCGGGGGAGA 1161

Qy    191 GCGGGTGCTAGGCGCGCTGCTGAGTAAGGCGCGTGCGCCGCCACGGCGGGTCGGTGAC 250
          ||| | |||||||||| |||| || || || ||  || | ||||||||||||||
Db   1162 CATTCTCCTGAGCAGCCTCATCAGAAAGAAGCTGCTTCCCGAGCCCTCTCTGCTCATCAC 1221

Qy    251 CACGCGCGCCGCCGCCCCCCGGGAGGCTGCGAGGGCGGCGAGGCGGCCTCGAGGCGCCGCGA 310
          ||||| | |||||| |||| |  | ||  |||||
Db    251 ... ... ...
```

RESULT 5
US-10-132-967-6
; Sequence 6, Application US/10132967
; Publication No. US20030170841A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THE
; FILE REFERENCE: 07334-311001
; CURRENT APPLICATION NUMBER: US/10/132,967
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: US/09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-132-967-6

Query Match     8.4%;    Score 226.4;    DB 12;    Length 3102;
Best Local Similarity  56.3%;   Pred. No. 6.5e-46;
Matches  466;   Conservative  0;   Mismatches 356;   Indels 6;   Gaps 2;

```
Qy     14 GCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCGACCAGTGCCCGACCGCGGCCGCCGTGCCGCA 73
          |||||||||||||||| |||| ||| ||||||||||||||||||||| || ||||
Db    801 GAGGAGCCTGGGGACGGGGACCTGATCATGAGCGTGCGCGGGGACCCCGACCGAACCACCACA 860

Qy     74 GATGCTGGCCGAGAAAACCCTCCAGAATCCTCTTCCCTATGGACCGGCTGC---C 130
          ||||||||||||||| || ||||||||||||||||| ||||
Db    861 GATGCTGGAGAAAACCCTCCAGAATCCTCTTCCTCTATGGACCGGCTTCGATGAGCTGCAAGG 920
```

```
Db   1222 CACGAGACCTGTGGCCCTGGAGAAACTGCAGCACTTGCTGCTGGACCATCTCCGCATGTGGA 1281
Qy    311 GGTGCGGCGGTTCTCCGACAAGGACACAAGAGAAGTATTTCTACAAGTTCTTCCGGGATGA 370
          || ||||| |||| ||||||||||| || || |||  ||| |||| ||  | |||
Db   1282 GATCCTGGGTTTCTCCGAGGCCCAAAGAAGTACTTCTTCAAGTACTTCTTCTGTGA 1341

Qy    371 GAGGAGGGCCGAGCGCGCCACCGCTTCGTGGAAGGAACGAGACGCGTGTCGGCGTG 430
          ||||||||||| ||| ||| |||| ||||||||||||| ||| || ||  ||||||||
Db   1342 GGCCCAAGCCAGCAGGCGCAGCCTTCAGTCGTGATTCAGGAGAACGAGGCCCTCTTCACCATGTG 1401

Qy    431 CTTCGTGCCCTGTGTGGTCGGATGGTGTGCACCGTGCTGCGCCAGCAGTGTGGAGCTCGG 490
          |||||||| ||| ||||| ||||||||| ||||| || || || ||||||  |||||||
Db   1402 CTTCATCCCCCGGTCTGCTGATCGTGTGCACTGGACTGAAACAGCAGATGGAGAGTGG 1461

Qy    491 TCGGGACCTGTCGCCAGTCCAAGACCAACCAGTGCAGTGGACCTGTTTCATCACCAG 550
          ||||||||||||||||| |||||||||||||||||||||||||||| ||
Db   1462 CAAGAGCCTTGCCCAGACATCCAGAGCCACCACACCCGGGTGACGTTACGTCTTCTTCCTTCCAG 1522

Qy    551 CGTTCTGAGCTGGCTTCGGTAGCCGACGGGCCCCGTTGCAGGGGCGACCTGCGAATC 610
          || | | |||||  |||
Db   1522 TTTGCTGCAGCCCCGGGGAGGAGGCGCAGGCACGGCGGCTCGCGCCACCCCGGGGCT 1581

Qy    611 GTGCCCCTGGCCCCGGCAGGCGTCTCCGACGCAGCAGGGCCAGTTGCCGAGGAAGGAAACT 670
          ||||| | || || || ||| |||| |||||||||||||
Db   1582 CTGCTCTTTGGCTGCAGATGGAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCT 1641

Qy    671 GGAGCAACTGAGCTTCGTGCTCCAAAGTGCGAGCGCGTGTTCTCAGCAAAAGGAGGTT 730
          |||||||||||||||||||| ||||| || | |||||||| |  |||||||||||||
Db   1642 CAGGAATCATGACTGCAGAAGGCGGATGT---GTCTGCTTTCCTGAGGATGAACCTGTT 1698

Qy    731 GCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTT 790
          ||||||| |||||| ||| || ||||||| |||| ||||||| | ||||||||||||||
Db   1699 CCAAAAGGAAGTGGACTGGCAGAAGTTCTACAGCTTCTACAGCTTATCATCCACCATGACTTCCAGGAGTT 1758

Qy    791 CCTCGGCCACTGTCCTACCTGCTGAGGACGACGGCGGGGTGCCCAGGAC 838
          |||| | ||||| | || |||||| ||
Db   1759 CTTTGCGCGCCCATGTACTACCTGCTGGAAGAGAGGAAAAGGAAGGAAGGAAGGAC 1806
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-127-516-6

Qy   Query Match           8.4%;  Score 226.4;  DB 13;  Length 3102;
Db   Best Local Similarity 56.3%;  Pred. No. 6.9e-46;
     Matches 466;  Conservative 0;  Mismatches 356;  Indels 6;  Gaps 2;

Qy     14 GGCGAGCCTGGCTGACCTGATCCTGACGAGTGCCCCGACCGGCGGCCCGGTGCCGCA  73
          ||||||||||||| ||||| |||||||||||  ||| |||||||||  ||||||||
Db    801 GAGGAGCCTGGGGACCTGATCATGAGCTGTCTCCCAAACCCACCCATCCACCACAA  860

Qy     74 GATGCTGGCCAGCCGCAGCGGCTGCTTCATCCTGGACCAGCGGCCGGCGACGAGCTGC--C  130
          |||| ||  |  ||| | || ||   ||||||||| |||||||||||||||||| |||
Db    861 GATCGTGAGAAAACCCTCCTGCTTCCTGATGGACGACGAGGCGCGAGCTCGATGAGTCAACG  920

Qy    131 GGGGCTGGGGGGCCCCGAGGCCGGCCTTCGTGAAGGAACGAGAGCGTGTTCGGCGGCTGTG  190
          |||||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db    921 TGCCTTTGACGAGCACATAGGACCGCTCCTCGCACTGACTGGCAGAGGCCGAGCGGGGAGA  980

Qy    191 GCGGGTGCTAGGGCGGCTGCTGAGTAAGGCGTGCCGCAGCAGCTGGAGCTGGTGAC  250
          ||||||||| ||||||||||||||||||   || |||||||||||||||||||||
Db    981 CATTCTCCTGAGCAGCCTCATCAGAAAGAGTGCTTCCCGAGGCTCTCTGCTCATCAC  1040

Qy    251 CACGGCGCGCGCCCCCGGAGGCTGCAGGGCGCCCGTGTTCCCGCAGTGCGCCGA  310
          ||| |||||||||||||||||||||| |||||||||||||||||||||||||
Db   1041 CACGAGACCTGTGGCCATCTCGGACCATCTCGGCATGTGGA  1100

Qy    311 GGTGGCGGGCTTCTCGACAAGGACAAGAAGAGTATTCTACAAGTTCTTCCGGGATGA  370
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1101 GATCCTGGGTTTCTCGAGGCCAAAAGGAAAGACTACTTCTTCAGTACTTCTTCTGATGA  1160

Qy    371 GAGGAGGGCCGAGCGGCTTCTCGACGCCCCGGTGCAGGGCGACCTGCGGAACT  430
          ||||| |||||||||||||||||||||||||||||||||||||||||||||
Db   1161 GAGGAGGGGCCGAGCGGCGTTCCTGGAGGAGAAGAGACGCGTGTTCGGCGGCTGTG  1220

Qy    431 CTTCGTGCTGGATCGTGCGGACCGTGCCCGACCTGGAGCTCGG  490
          ||| ||||||||||||||||||||||||||||||||||||||
Db   1221 CTTCATCCCCTGGTCGTGGATCGTGCGGACCGTGCGAAACAGCGAGATGGAGAGTGG  1280

Qy    491 TCGGGACCTGTCGCGCAGACAGTCCAAGACCACCAGTCAGTGTCTGCTTTCATCACCAG  550
          ||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db   1281 CAAGAGCCTTGCCCAGCATCTAAGACCCACCACCACGCGGTGTACGTCTTCCTTTCCAG  1340

Qy    551 CGTTCTGAGCTCGGCTCCGGACCCCGGTCAGGCGGGCCAGGGCGGGGTGCCCAATCT  610
          ||| || ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1341 TTTGCTGCAGCCCCCGGGAGGGAGGCACGGAGGCAGGAGCACGGCCCTCGCCCACCTCTGGGGCT  1400

Qy    611 GTGCCGCCTGGCCCGGAGCCCGAGGGCGTTGCCGAGAAGGAACT  670
          ||||||||||||||||||||||||||| |||||||||||||
Db   1401 CTGCTCTTTGGCTGCAGATCGGAATCTGAACCAGAAAAACCTGTTGGAGAGTCCGACC  1460

Qy    671 GGAGCAACTGGAGCTTCGTGGCTCCAAAGTCGAGAGGCGGATGT--GTCTGCTTTCCTGAGGATGAACCTGT  730
          ||||||||||||||||||||||||||||||||||| ||||||    |||||||||||||||||||||
Db   1461 CAGGAATCATGGACTGCAGAGGCGGATGT--GTCTGCTTGTTCCTGCACCAGAGCTTCAGGAGT  1517

Qy    731 GCCGGGCGTGCTGAGACAGAGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTT  790
          |||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db   1518 CCAAAAGAAGTGACTGCCGAGAAGTTCTACAGCTTCATCCACCATGACTTTTCAGGAGTT  1577

Qy    791 CCTCGGCGACTGTCCTACCTGCTGTGAGGACGCGGGGTGCCCAGGAC  838
          ||| |||||||||||||||||| |||||||||||||||||||
Db   1578 CTTTGCCGCCATGGCCATGTACTACCTGCTGTGAAGAGGAAAGGAAGGAGGAC  1625
```

RESULT 7
US-10-027-629-6
; Sequence 6, Application US/10027629
; Publication No. US20020197660A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  APPLICANT: Manji, Gulam A.
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN

; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-341001
; CURRENT APPLICATION NUMBER: US/10/027,629
; CURRENT FILING DATE: 2001-12-20
; PRIOR APPLICATION NUMBER: US 09/964,955
; PRIOR FILING DATE: 2001-09-26
; PRIOR APPLICATION NUMBER: US 09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 6
; LENGTH: 3102
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-027-629-6

Query Match        8.4%;     Score 226.4;    DB 13;    Length 3102;
Best Local Similarity   56.3%;    Pred. No. 6.5e-46;
Matches  466;   Conservative   0;   Mismatches  356;   Indels  6;   Gaps  2;

```
Qy     14  GCGCAGCCTGCTGACCTGATCCTGACCAGTGCCCGACCGGCGGCGGGTGCCGCA    73
           ||||||||||||  ||  |  |  |||||       ||||||  ||| |||| |||
Db    801  GAGGAGCCTGGGGGACCTGATCATGAGCTGCTGCCCCGACCCACCATCCACAA   860

Qy     74  GATGCTGGCCCAGCCAGCGGCTGCTCTTCATCCTGACAGCGGCGGACGAGCTGC---C  130
           ||||||||| | ||   |  |  |||   |||   |||||||||| ||||| |  |
Db    861  GATCGTGAGAAAACCCTCCAGAATCCTCTTCCTCATGGACGGCTTCGATGAGCTGCAAGG  920

Qy    131  GGCGCTGGGGGCCCCGAGGCGGCCCTTCGAGGCGGCGGCGCGC                  190
           ||||   |||||  || | |  |  |||||||||||||||| |
Db    921  TGCCTTTGACGAGCACATAGGACCCGCTGCGACTGACTGGAGCGGCGGGAGA        980

Qy    191  GCGGGTGCTAGGCGGGCGTGCTGAGTAAGGCCGTGCCCACGGCCCTCCTGCTGTGAC    250
           ||||||   |  ||  |  |  |||||||| |  |  |||||||||||||   |  ||
Db    981  CATTCTCCTGAGCAGCCTCATCATCAGAAAGAAGCTGCTTCCCGAGGCCTCATCATCAC  1040

Qy    251  CACGCGCGCGCCGCGCCCCGGGAGGCTGCAGGGGCCGCGTGTTCCCGCAGTGCGCCGA   310
           |||||||||||||||||  ||||||||| | |||   || |  |||||| ||| ||||
Db   1041  CACGAGACCTGTGCCCTGGAGAAACTGCAGCACTTGCTGACCATCCTCGGCATGTGGA  1100

Qy    311  GGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAGTATTTCTACAAGTTCTTCCGGGATGA   370
           |||||||||||||   |||  || || |  || | || |||||||  |||  |||||||
Db   1101  GATCCTGGGTTTCTCCGAGGCCAAAAGGAAGGCCCAAAAGGAAAGAGTACTTCTTCAAGTACTTCTCTGATGA  1160

Qy    371  GAGGAGGCGCCAGCGGCGCTACCGCTTCGTGAAGGAAGGAACGAGACGCGTGTGGCGCGTG   430
           ||||| | |  ||   |||||  | |||| ||||||||| ||||  | |  |  |  |
Db   1161  GGGCCCAAGCGCAGGGGACCGTTCCGAGGCGCAGTTTGCCGACTGCTGACTGCACTGCACTTGCGCGGAGAGAGGAACT  1220

Qy    491  TCGGGACTGTGCGGCAGCTCAGTCCAAGACGTCAGTGTACCTGCCTTCGCCTGACTGCAGTGTC    550
```

---

RESULT 8
US-10-132-967-4
; Sequence 4, Application US/10132967
; Publication No. US20030170842A1
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THI
; FILE REFERENCE: 07334-311001
; CURRENT APPLICATION NUMBER: US/10/132,967
; CURRENT FILING DATE: 2002-04-26
; PRIOR APPLICATION NUMBER: US/09/653,901
; PRIOR FILING DATE: 2000-09-01
; PRIOR APPLICATION NUMBER: US 09/506,067
; PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 13
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
; LENGTH: 3857
; TYPE: DNA
; ORGANISM: Homo sapiens
US-10-132-967-4

Query Match        8.4%;     Score 226.4;    DB 12;    Length 3857;
Best Local Similarity   56.3%;    Pred. No. 6.9e-46;
Matches  466;   Conservative   0;   Mismatches  356;   Indels  6;   Gaps  2;

```
Qy     14  GCGCAGCCTGCTGACCTGATCCTGACCAGTGCCCGACCGGCGGCGGGTGCCGCA      73
           ||||||||||||  ||  |  |  |||||       ||||||  ||| |||| |||
Db    939  GAGGAGCCTGGGGGACCTGATCATGAGCTGCTGCCCCGACCCACCATCCACAA     998

Qy     74  GATGCTGGCCCAGCCAGCGGCTGCTCTTCATCCTGACAGCGGCGGACGAGCTGC---C    130
Db    999  GATCGTGAGAAAACCCTCCAGAATCCTCTTCCTCATGGACGGCTTCGATGAGCTGCAAGG   1058

Qy    131  GGCGCTGGGGGCCCCGAGGCGGCCCTTCGAGGCGGCGGCGCGC                   190
Db   1059  TGCCTTTGACGAGCACATAGGACCCGCTGCGACTGACTGGAGCGGCGGGAGA         1118

Qy    191  GCGGGTGCTAGGCGGGCGTGCTGAGTAAGGCCGTGCCCACGGCCCTCCTGCTGTGAC    250
Db   1119  CATTCTCCTGAGCAGCCTCATCATCAGAAAGAAGCTGCTTCCCGAGGCCTCATCATCAC   1178

Qy    251  CACGCGCGCGCCGCGCCCCGGGAGGCTGCAGGGGCCGCGTGTTCCCGCAGTGCGCCGA    310
Db   1179  CACGAGACCTGTGCCCTGGAGAAACTGCAGCACTTGCTGACCATCCTCGGCATGTGGA   1238

Qy    311  GGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAGTATTTCTACAAGTTCTTCCGGGATGA    370
Db   1239  GATCCTGGGTTTCTCCGAGGCCAAAAGGAAGGCCCAAAAGGAAAGAGTACTTCTTCAAGTACTTCTCTGATGA   1298

Qy    371  GAGGAGGCGCCAGCGGCGCTACCGCTTCGTGAAGGAAGGAACGAGACGCGTGTGGCGCGTG    430
Db   1299  GGGCCCAAGCGCAGGGGACCGTTCCTGTGTGTGCGATCGATTCAGGAGAGAACGAGGTCCTCG   1358

Qy    431  CTTCGTGCCCTTCGTGTGCGCAGCTCGGGGCGGTGGGAGAGCAGCAGTGGAGCTGGAGTGG    490
Db   1359  CTTCATCCCCGGTCTTGCTGCTGGATCGTGTCGCGATGGGACTGCACTGACAGCAGTAGTGG   1418

Qy    491  TCGGGACTGTGCGGCAGCTCAGTCCAAGACGTCAGTGTACCACCAG     550
Db   1419  CAAGAGCCTTGCCCAGACATCTAAGACACCAAGCGCGGGTACGTTTCCTTCCAG   1478

Qy    551  CGTTCTGAGCTCGGCCGGGGAGGGAGGGCTCGACTTGCAGGCGGCCAATCT     610
Db   1479  TTTGCTGCAGCCCCGGGGAGGGAGGAGCCAGGAGCCACCCTCGGGGGCT   1538

Qy    611  GTGCCGCCTGGCCCGCGGGAGGGCGTCCTCGGACGCAGCGGCCAGTTTGCCGAGAAGGAACT    670
```

```
                |||    ||||    || ||  ||      |  |     | |||||  |||   || ||
Db        1539 CTGCTCTTTGGCTGCAGATGGAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCT 1598

Qy         671 GGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCT 730
                  | | |  |  || |      |  | ||   | |    || || || |  ||   | |
Db        1599 CAGGAATCATGGACTGCAGAAGGCGGATGT---GTCTGCTTTCCTGAGGATGAACCTGTT 1655

Qy         731 GCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTT 790
                |     |   ||||    |||     ||||   |||||| ||  ||  |||||||||||
Db        1656 CCAAAAGGAAGTGGACTGCGAGAAGTTCTACAGCTTCATCCACATGACTTTCCAGGAGTT 1715

Qy         791 CCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGAC 838
               | | || ||  ||| |||||||||||| ||  |    ||     |||||
Db        1716 CTTTGCCGCCATGTACTACCTGCTGGAAGAGGAAAAGGAAGGAAGGAC 1763

RESULT 9
US-10-127-516-4
; Sequence 4, Application US/10127516
; Publication No. US20020187922A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  APPLICANT:  Manji, Gulam A.
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-370001
;  CURRENT APPLICATION NUMBER: US/10/127,516
;  CURRENT FILING DATE:  2002-07-23
;  PRIOR APPLICATION NUMBER: US 10/027,629
;  PRIOR FILING DATE: 2001-12-20
;  PRIOR APPLICATION NUMBER: US 09/964,955
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 09/653,901
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: US 09/506, 067
;  PRIOR FILING DATE: 2000-02-17
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 3857
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (139)...(3240)
US-10-127-516-4

  Query Match              8.4%;  Score 226.4;  DB 13;  Length 3857;
  Best Local Similarity   56.3%;  Pred. No. 6.9e-46;
  Matches  466;  Conservative    0;  Mismatches  356;  Indels    6;  Gaps    2;

Qy          14 GCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCGCGCCGGTGCCGCA 73
               | | |||||||  ||||||||| || ||  |   ||||||||||    | | | |  |   |
Db         939 GAGGAGCCTGGGGGACCTGATCATGAGCTGCTGCCCCGACCCAAACCCACCCATCCACAA 998

Qy          74 GATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGGACGAGCTGC---C 130
               |||  ||    | ||     |  | ||||| | || |||||||   || |||||||    
Db         999 GATCGTGAGAAAACCCTCCAGAATCCTCTTCCTCATGGACGGCTTCGATGAGCTGCAAGG 1058

Qy         131 GGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGC 190
                ||  | |  |  | |   |   ||| |||||| |||       |||| | | | || | 
Db        1059 TGCCTTTGACGAGCACATAGGACCGCTCTGCACTGACTGGCAGAAGGCCGAGCGGGGAGA 1118

Qy         191 GCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCCACGGCCCTCCTGCTGGTGAC 250
                    | ||  || | ||  | || |||  |||||| |||  ||||   |||||  | ||
Db        1119 CATTCTCCTGAGCAGCCTCATCAGAAAGAAGCTGCTTCCCGAGGCCTCTCTGCTCATCAC 1178

Qy         251 CACGCGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGCAGTGCGCCGA 310
               |||| |  | |  ||||  | ||  ||||||  |   |||      || | |   |  ||
Db        1179 CACGAGACCTGTGGCCCTGGAGAAACTGCAGCACTTGCTGGACCATCCTCGGCATGTGGA 1238

Qy         311 GGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGGATGA 370
               | | |  || ||||||||     | | |||  |||| ||||| ||||| ||||   |||||
Db        1239 GATCCTGGGTTTCTCCGAGGCCAAAAGGAAAGAGTACTTCTTCAAGTACTTCTCTGATGA 1298

Qy         371 GAGGAGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTG 430
               |      |||  |   |||| | |  |  |  ||||||||||||   || ||| |  ||||
Db        1299 GGCCCAAGCCAGGGCAGCCTTCAGTCTGATTCAGGAGAACGAGGTCCTCTTCACCATGTG 1358

Qy         431 CTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGCTCGG 490
               |||| | ||| | || |||||||||||||||||| |  |||   ||||||| |||||   ||
Db        1359 CTTCATCCCCCTGGTCTGCTGGATCGTGTGCACTGGACTGAAACAGCAGATGGAGAGTGG 1418

Qy         491 TCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCATCACCAG 550
                  |  |||  | |  || || ||||||||||||  | |||||| |  | ||  |||
Db        1419 CAAGAGCCTTGCCCAGACATCTAAGACCACCACCGCGGTGTACGTCTTCTTCCTTTCCAG 1478

Qy         551 CGTTCTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCT 610
                 | |||    |       |    | | | || | |  |    | | |||||  |    ||
Db        1479 TTTGCTGCAGCCCCGGGGAGGGAGCCAGGAGCACGGCCTCTGCGCCCACCTCTGGGGGCT 1538

Qy         611 GTGCCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGCGCAGTTTGCCGAGAAGGAACT 670
                |||    ||||    || ||  ||      |  |     | |||||  |||   || ||
Db        1539 CTGCTCTTTGGCTGCAGATGGAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCT 1598

Qy         671 GGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCT 730
                  | | |  |  || |      |  | ||   | |    || || || |  ||   | |
Db        1599 CAGGAATCATGGACTGCAGAAGGCGGATGT---GTCTGCTTTCCTGAGGATGAACCTGTT 1655

Qy         731 GCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTT 790
                |     |   ||||    |||     ||||   |||||| ||  ||  |||||||||||
Db        1656 CCAAAAGGAAGTGGACTGCGAGAAGTTCTACAGCTTCATCCACATGACTTTCCAGGAGTT 1715

Qy         791 CCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGAC 838
               | | || ||  ||| |||||||||||| ||  |    ||     |||||
Db        1716 CTTTGCCGCCATGTACTACCTGCTGGAAGAGGAAAAGGAAGGAAGGAC 1763

RESULT 10
US-10-027-629-4
; Sequence 4, Application US/10027629
; Publication No. US20020197660A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  APPLICANT:  Manji, Gulam A.
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
;  TITLE OF INVENTION:  PROTEIN FAMILY AND USES THEREOF
;  FILE REFERENCE: 07334-341001
;  CURRENT APPLICATION NUMBER: US/10/027,629
;  CURRENT FILING DATE:  2001-12-20
;  PRIOR APPLICATION NUMBER: US 09/964,955
;  PRIOR FILING DATE: 2001-09-26
;  PRIOR APPLICATION NUMBER: US 09/653,901
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: US 09/506,067
;  PRIOR FILING DATE: 2000-02-17
;  NUMBER OF SEQ ID NOS: 17
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 4
;   LENGTH: 3857
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (139)...(3240)
US-10-027-629-4

  Query Match              8.4%;  Score 226.4;  DB 13;  Length 3857;
  Best Local Similarity   56.3%;  Pred. No. 6.9e-46;
  Matches  466;  Conservative    0;  Mismatches  356;  Indels    6;  Gaps    2;

Qy          14 GCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCGCGCCGGTGCCGCA 73
               | | |||||||  ||||||||| || ||  |   ||||||||||    | | | |  |   |
Db         939 GAGGAGCCTGGGGGACCTGATCATGAGCTGCTGCCCCGACCCAAACCCACCCATCCACAA 998

Qy          74 GATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGGACGAGCTGC---C 130
```

Best Local Similarity: 55.0%;   Pred. No. 1.5e-39;
Matches 437;   Conservative   0;   Mismatches 351;   Indels   6;   Gaps   2;

```
Qy    189  GCGGGGGTGCTAGGCGGCTGCTGAGTAAGGCGGCTAAGGCCCTTCCTGCTTCGGTCGGTG  248
           ||||||||| | ||||| | |||||   ||||||  | |||||||||  || |||||||||
Db    247  GAGCTGCTTCTTAACAGCTAAATTCGGAGAAGCTGCTCCCTGAGCTATCTTTGCTCATC   306

Qy    249  ACCACGCGGCGGCCGCCGCCCCGGAGGCTGCTGCCAGTGCGGCCGCCTCCTGGTCCCCGAGTGCGCC  308
           |||||| ||||||||||||| | ||||||||||||||  ||||| ||||||   | |||||| || |
Db    307  ACCACGCGGCCCACGGCTTTGGAGAAGCTCACCGTCGCTGGAGCACCCCAGGCATGTG   366

Qy    309  GAGGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGAGTATTTCTACAAGTTCTTCGGGAT  368
           | || | || | |||||||  | |||||||||  ||||||  |||||| | | || ||||||
Db    367  GAGATCCTGGCTTCTCTGAGCGCAGAAAGGAAGGAAGGACAGTATTTCTACAAGTATTTCCACAAT   426

Qy    369  GAGAGGAGGGCCGAGCGCGCGCTACCGCTTCGTGATGCTGGATGA  428
           |   |||  | |||||| | | |||  |||||||||||||||||||
Db    427  GCAGAGCAGGCGGGGCCAAGTCTTCAATTACGTGAGGGACAACGAGCCTCTCTTCCACCATG   486

Qy    429  TGCTTCGTCCCTTCGTGCTGGATCGTGCGCAGCGTCGTGCGCCAGCAGTGGAGCTC  488
           |||||||||  ||| | |||||||| ||| |||||||   ||| |||||||||||||
Db    487  TGCTTCGTCCCCCTGGTGCTGGTGGGTGGTGTACCTGCCTGCCTCCAGCAGCTGGAGGGT   546

Qy    489  GGTGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTCATCACC  548
           ||||||| || |  || ||||| || | |||||  ||| | || || ||||||||||||
Db    547  GGGGGCGCTGTTGAGACGTCCAGGACGTCCAGGACACTGCAGTGTACACTGCTACCTCTG   606

Qy    549  AGGGTTCTGAGCTCGGCTCCGGTTGCGAGGCTCCGGACCTGCGCAAT  608
           ||  | | ||| |  | ||| |  |  |||||   ||  ||| ||||||
Db    607  AGTCTGATGCAACCCAAGCCGGGGCCCGCCTCCAGCCCCCAACCAGAGAGG   666

Qy    609  CGTTGCCGCCTGGCCTCGGCTCCGGTTGCAGGGCGACCTGCAGAAGGAA  668
           | |||| || |||| ||| ||  |||||  | |||  |||||||||||||
Db    667  TTGTGCTCCTTGGCGGCAGATGGCTTCTGGAATCAGAAAATCCTATTTGAGGAGGACAGGAC   726

Qy    669  CTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCCGTTTCTCAGCAAAAAGGAG  728
           |||||| ||||||| ||| |   | |  ||||  | |||  |  |||  || |||||  |||
Db    727  CTCCGGAAGCAGCACGCGCCT---AGACGGGGAAGCAGCGTCTCTGCCTTCCTCCAACATGAACATC   783

Qy    729  CTGCCGGGGGCTGCTGGAGACAGAGGTTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAG  788
           |||| ||||| |||||| | |||| |||||| ||| || |||||| | |||||||||||||||
Db    784  TTCCAGAAGACATCAACTGTGAGGAGTACTACAGCTTCATCCACTTGAGTTTCCAGGAA   843

Qy    789  TTCCTGCGGGGCACTGTCCTACCTGCTGGAGACAGCGGCGGGGGTGCCCGGCGGCTGGC  848
           |||| |  ||  |   ||| | || | |||||||   || | || |||  | |||||| |
Db    844  TTCTTTGCGCTATGTGACTATATATCCTGACGCAGAGGGAGGGGGCAGGCCCAGGACCAG   903

Qy    849  GGCGTTGGACACTCCTGCCTGGGAGGACGGGGCCCAGCCGCCACTTGGTG---CTCACC  905
           || ||||||| | ||  || |||||||||  |  ||| ||  ||  ||||||    |||||
Db    904  GACGTGACCAGGCTGTTGACCGAGTACGCGTTTTCTGAAAGGAGCTTCCTGCGCACTCACC   963

Qy    906  ACGGGCCTTCCTCTTCGGACTGCTCGAGGCCGGATGCGCGACATCGAGCGGCCACTTC  965
           || ||  || || |||  ||||||| || |  ||| ||||||  || ||||||||||||
Db    964  AGCCCCTTCCTCTGTTTGGACTCCTGAACGAGGACCCACCTGGAGAAGAGTCTC  1023

Qy    966  GGCTGCATGGTTTC  979
           ||||| || ||| ||
Db   1024  TGCTGGAAGGTCTC  1037
```

RESULT 12
US-10-132-967-3
; Sequence 3, Application US/10132967
; Publication No. US20030170841A1
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THER
;  FILE REFERENCE: 07334-311001
;  CURRENT APPLICATION NUMBER: US/10/132,967
;  CURRENT FILING DATE: 2002-04-26
;  PRIOR APPLICATION NUMBER: US/09/653,901
;  PRIOR FILING DATE: 2000-09-01
;  PRIOR APPLICATION NUMBER: US 09/506,067

---

```
Db    999  GATCGTGAGAAAACCCTCCAGAATCCTCTTCCTCCTCATGCATGGACGGCTTCGAGGCGGCGAGCGGCGC  1058
           || ||  |   ||     | ||| ||| |||||||| | ||   |||||||| |||||||| |||||||||
Qy    131  GGCGCTGGGGGTGCCCCCGAGCGCGCCCTGCACAGACCCCCTTCGAGGCGGCGAGCGGCGC   190

Db   1059  TGCCTTTGACAGCACATAGACCGTCTGCACTGCAGAGGGCGAGCGGGGGAGA  1118
           ||||| ||||| |||  ||| ||| |||||| |  ||| |||| ||| |||||
Qy    191  GCGGGGTGCTAGGCGGCTGCTGAGTAAGGCGGCTGCTGCCAGTGCGGCC   250

Db   1119  CATTCTCCTGAGCAGCCTCATCAGAAAGAAGCTGCTTCTCCCGAGGCCTCTGCCCATCAC  1178
           ||||||| || |||||| ||| |||| |||| |||| ||||| | |||   ||| |||||
Qy    251  CACGCGGCCGCCGCCCCGGAGGCTGCAGGCCCGCCTGTGTTCCCCGAGTGCGCCGA   310

Db   1179  CACGAGACCTGTGGCCCTGGAGAAACTGCTGCACTGCTTGCTGACCATCCTCGGCATGTGGA  1238
           ||| ||| ||| |||||||||| || |||||||| ||||  ||| |||  ||| |||||||
Qy    311  GAGGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGAGTATTTCTACAAGTTCTTCGGGATGA   370

Db   1239  GATCCTGGTTTTCTCTGAGCGCAAAAGAAGAGTACTTCTTCAAGTACTTCTCTGATGA  1298
           || | || |||||||| |||   |||||||| | || ||||||| |||||  ||| ||
Qy    371  GGCGCGGGCTTCTCCGGCCTACCGCTTCGTGATGCTGGATCGTGCGCAGCGTCGTGCGCCAGCAGC   430

Db   1299  GGCCCAAGCCAGGCAGCCTCAGTCTGATTCAGGAGACGAGGTGCTCCTTCACCATGTG  1358
           ||| ||| ||||| |||| || |   |||| |||| |||  |  | ||| | ||||| |
Qy    431  GGCGGGCCTACCGCGCGCCTACCGCGTCGTGCGCCAGCAGCGTCCAGCAGCTGGAGTCGG   490

Db   1359  CCTCATCCCCGTCGTCTGCTGCTGTCTGCATCGGACTCTGTCTGACTCGGACGATGGAGAGTGG  1418
           || ||||  |||| |||| ||| |  || || ||| |  | |||| |||  | |  |||||||||
Qy    491  CCTCCGACCGTGTGCTGCGCCACCACGTCGTCTCGTTACCTCCACTGGACGAG   550

Db   1419  CAAGAGCCTTGCCCAGACACATCCAAGCACCACCGGGTACGTCTTTCCTTCCTTTCCAG  1478
           ||||| ||| |  ||| |||| ||| || | ||  || | | ||||||||| |  ||||
Qy    551  TCGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTCATCACCAG   610

Db   1479  TTTGCTGCAGCCCGGCCCGGGGGAGGGACGGCCAGGAGCCTCGCGCCCACCTCTGGGGGCT  1538
           | |||| || |  | |||| || | ||| ||| |||||||| || || |||||||| ||
Qy    611  CGTTGCCGCCTGGCCTCGGCTCCGGTTGCAGGGCGACCTGCAGAAGGAACT   670

Db   1539  CTGGTCTTTGGCTGCGGCAGTGGAATCTGGAACCAGAGAAATCCTGTTTGAGGAGGTCCGACCT  1598
           ||||  | |  ||| || | | ||||||||| ||||| | ||| | ||||||||||| ||| ||
Qy    671  GGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCCGTTTCTCAGCAAAAAGGAGC   730

Db   1599  CAGGAGAATCATGCAGCTGACTGCAGAGAGGCCGATGT---GTCTGCTTTCCTGAGGATGAACCTGTT  1655
           |||||| | | |||| | ||| |||| ||  |||| ||   |||| ||||||| || |||| ||||
Qy    731  GCCGGGCGTGCTGGAGACAGAGGTTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTT   790

Db   1656  CCAAAAGGAAGTGGACTGCGCGAGAAGTTCTACCGAGTTCATCCACATGACTTCCAGGGAGTT  1715
           || |||||| || | | ||| ||| ||||| | ||| ||||||||  |  ||||||||| |||
Qy    791  CCTGCGGGGCACTGTCCTACCTGCTGGAGACGGCGGCGGGGGTGCCCGGCGGCTGGC   838

Db   1716  CTTTGCGCGACCATGCAGTACTACCTGGTGGAGAAGAGGAAAAGGAAGGAAGGAAGGAGGAAGGAC  1763
           |  ||| |  |||||| |||| ||| || |||| ||| |||||| || || |  || || |||
Qy    839  GGCGTTGGACACTCCTGCCTGGGAGGACGGGGCCCAGC   763
```

RESULT 11
US-09-833-381-1976
; Sequence 1976, Application US/09833381
; Patent No. US20020132090A1
; GENERAL INFORMATION:
;  APPLICANT: Robison, Keith E.
;  TITLE OF INVENTION: No. US2002013209OAlel Nucleic Acid and Protein Homologs
;  FILE REFERENCE: 5800-119
;  CURRENT APPLICATION NUMBER: US/09/833,381
;  CURRENT FILING DATE: 2001-04-11
;  PRIOR APPLICATION NUMBER: 09/516,448
;  PRIOR FILING DATE: 2000-02-29
;  NUMBER OF SEQ ID NOS: 2050
;  SOFTWARE: FastSEQ for Windows Version 3.0
; SEQ ID NO 1976
;  LENGTH: 1650
;  TYPE: DNA
;  ORGANISM: Homo sapiens
US-09-833-381-1976

Query Match       7.5%;   Score 200.4;   DB 10;   Length: 1650;

```
;  PRIOR FILING DATE: 2000-02-17
;  NUMBER OF SEQ ID NOS: 13
;  SOFTWARE: FastSEQ for Windows Version 4.0
;  SEQ ID NO 3
;  LENGTH: 3099
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (1)...(3099)
US-10-132-967-3

Query Match          6.7%;  Score 179.2;  DB 12;  Length 3099;
Best Local Similarity  54.0%;  Pred. No. 2.9e-34;
Matches  445;  Conservative  0;  Mismatches  358;  Indels  21;  Gaps  3;
```

RESULT 13
US-10-127-516-3
; Sequence 3, Application US/10127516
; Publication No. US20020187922A1
; GENERAL INFORMATION:
;    APPLICANT: Bertin, John
;    APPLICANT: Manji, Gulam A.
;    TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
;    TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;    FILE REFERENCE: 07334-370001
;    CURRENT APPLICATION NUMBER: US/10/127,516
;    CURRENT FILING DATE: 2002-07-23
;    PRIOR APPLICATION NUMBER: US 10/027,629
;    PRIOR FILING DATE: 2001-12-20
;    PRIOR APPLICATION NUMBER: US 09/964,955
;    PRIOR FILING DATE: 2001-09-26
;    PRIOR APPLICATION NUMBER: US 09/653,901
;    PRIOR FILING DATE: 2000-09-01
;    PRIOR APPLICATION NUMBER: US 09/506,067
;    PRIOR FILING DATE: 2000-02-17
; NUMBER OF SEQ ID NOS: 17
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;    LENGTH: 3099
;    TYPE: DNA
;    ORGANISM: Homo sapiens
US-10-127-516-3

Query Match          6.7%;  Score 179.2;  DB 13;  Length 3099;
Best Local Similarity  54.0%;  Pred. No. 2.9e-34;
Matches  445;  Conservative  0;  Mismatches  358;  Indels  21;  Gaps  3;

[Sequence alignment data follows in two mirrored columns]

```
Db   1260  GACGCTGAGCCTCCNGGCCGCGCAGGGCCGTGTGGGCGCAGACGTCCGTGCTTCACCGAGA 1319
                  ||||||  ||||    |||||||
Qy    665  GGAACTGGAGCAACTGGAGC:TCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAA 724
                  ||||||  ||||||
Db   1320  GGATCTGGAAAGGCTCGGGG:GCGAGAGTCCGACCTCC---GTCTGTTCCTGGACGGGAGA 1376
                  ||||||  |||||
Qy    725  GGAGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCA 784
                  ||||||  |||||||
Db   1377  CATCCTCCGCCAGGACAGAGTCTCCAAAGGCTGCTACTCCTTCATCCACCTCAGCTTCCA 1436
                  ||||||  |||||
Qy    785  GGAGTTCTCGGGCCACTGTCCTACCTGCTGGAGGACGGCGGGG 828
                  ||||||
Db   1437  GCAGTTTCTCACTGCCCTGTTCTACACCCTGGAGAAGGAGGAGG 1480
```

RESULT 14
US-10-027-629-3
; Sequence 3, Application US/10027629
; Publication No. US20020197660A1
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   APPLICANT: Manji, Gulam A.
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN
;   TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;   FILE REFERENCE: 07334-341001
;   CURRENT APPLICATION NUMBER: US/10/027,629
;   CURRENT FILING DATE: 2001-12-20
;   PRIOR APPLICATION NUMBER: US 09/964,955
;   PRIOR FILING DATE: 2001-09-26
;   PRIOR APPLICATION NUMBER: US 09/653,901
;   PRIOR FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: US 09/506,067
;   PRIOR FILING DATE: 2000-02-17
;   NUMBER OF SEQ ID NOS: 17
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 3
;   LENGTH: 3099
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-027-629-3

Query Match      6.7%;   Score 179.2;  DB 13;  Length 3099;
Best Local Similarity   54.0%;   Pred. No. 2.9e-34;
Matches  445;  Conservative   0;  Mismatches  358;  Indels  21;  Gaps  3;

```
Qy     8  GTGGGCGCGCAGCCTGGCTGACCTGATCCTGACCTGACCCCGACCGCCCGGCGCGCCGGT 67
                ||||||
Db   675  GGGCCCGTGCAGTTTGCAGAGCTGGTCTTCAGGGACTGGCCTGAATTGCAGGATGACAT 734
                ||||||
Qy    68  GCCGCAGATGCTGGCCCAGCCGCAGCGCGGCCGGCGGCTGCTCTTCATCCTGGACGCGGCGACGAGCT 127
                ||||||
Db   735  TCCACACATCCTAGCCCAAGCACGGAAAATCTTGTTCGTGATTGACGGCTTTGATTGAGCT 794
                ||||||
Qy   128  G---CCGGCGTGGGGGGGCGCCAGGGCCCGGAGGGCCGGAGTA:TTCTACAAGTACTCCCG 184
                ||||||
Db   735  GGGAGCGCCGCACCTGGGGCGCGATCGAGGGACTGCGCGGGACTGCGCGGGAGGAAGAAGAGCC 854
                ||||||
Qy   185  CGGCGGGCGGGTGCTAGGCGGGTGCTGAGTAGAGGCGGCTGGCTGCCCACGGGCCTCCTGCT 244
                ||||||
Db   855  GGTGCCGTCTCCTCTGGGAGTTTGCTGAACAGGTGATGTTACCCGCCCTGTTCCCCGAGTG 914
                ||||||
Qy   245  GGTGACCACGCGCGCGCGCCACCGGAGGGCTGCAGGGCCGCCTGTTCCCGCAGTG 304
                ||||||
Db   915  GGTCACCACGCGCGCCGCCAGGGCCCTGGCCGGAGGGCCGGATCCA 974
                ||||||
Qy   305  CGCCGAGGTGCGCGGCTTCTCCCACACAAGGACAAGAAGAAGTACTTCTACAAGTAC:TCCCG 364
                ||||||
Db   975  CAAAAGGTGGAGGGCTTCCTGGAGGGCGCCCATTCCTGGACACTTGG 1034
                ||||||
Qy   365  GGATGAGAGGAGGGCGAGCGCGCACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGC 424
                ||||||
Db  1035  AGACGAGGACCAAGCCATGCGTCCTTGAGCTAATGAGGAGCAACGCGCCCTGTTCCA 1094
```

RESULT 15
US-10-132-967-1
; Sequence 1, Application US/10132967
; Publication No. US20030170841A1
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE PYRIN DOMAIN PROTEIN FAMILY AND USES THE
;   FILE REFERENCE: 07334-311001
;   CURRENT APPLICATION NUMBER: US/10/132,967
;   CURRENT FILING DATE: 2002-04-26
;   PRIOR APPLICATION NUMBER: US/09/653,901
;   PRIOR FILING DATE: 2000-09-01
;   PRIOR APPLICATION NUMBER: US 09/506,067
;   PRIOR FILING DATE: 2000-02-17
;   NUMBER OF SEQ ID NOS: 13
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 1
;   LENGTH: 3431
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-10-132-967-1

Query Match      6.7%;   Score 179.2;  DB 12;  Length 3431;
Best Local Similarity   54.0%;   Pred. No. 3e-34;
Matches  445;  Conservative   0;  Mismatches  358;  Indels  21;  Gaps  3;

```
Qy     8  GTGGGCGCGCAGCCTGCTGACCTGATCCTGACCAGTGCCCGACCGCCCGGCGCGCCGGT 67
                ||||||
Db   752  GGGCCCGTGCAGTTTGCAGAGCTGGTCTTCAGGGACTGGCCTGAATTGCAGGATGACAT 811
                ||||||
Qy    68  GCCGCAGATGCTGGCCCAGCCGCAGCGCGGCGGCCGCGTTGATGAGCT 127
                ||||||
Db   812  TCCACACATCCTAGCCCAAGCACGGAAAATCTTGTTCGTGATTGACGGCTTTGATGAGCT 871
                ||||||
Qy   128  G---CCGGCGTGGGGGGGCGCCAGGGCCCCTTCGAGGCGGCGAG 184
                ||||||
Db   872  GGGAGCCGCACCTGGGGCGCGTGATCGAGGACATCTGCGGGGACTGGGAGGAAGAAGAGCC 931
                ||||||
Qy   185  CGGCGGGCGGGTGCTAGGCGGGTGCTGAGTAAGGCGGCGTGCCCACGGGCCCTCCTGCT 244
                ||||||
Db   932  GGTGCCCGTCTCCTCTGGGGAGTTTGCTGAACAGGTGATGTTACCCGCCCTGTTCCCCGAGTG 991
                ||||||
Qy   245  GGTGACCACGCGCGCGCGGGAGGCTGCAGGGCCGCCTGTTCCCCGCAGTG 304
```

Sat Oct  4 14:14:49 2003

```
Db      992 GGTCACCACGGCGCGCCCAGGGCCCTGAGGGACTCCGGATCCTCGGCGGAGGAGCCGATCTA 1051
             |||  ||||||||||   ||   |||   |  ||                           |
Qy      305 CGCCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAGAAGTATTTCTACAAGTCCTTCCG 364
             |     ||||  ||||||||   |||      ||  ||    |  ||||
Db     1052 CATAAGGGTGGAGGGCTTCCTGAGGAGGACCAAGAGAGGGCTATTTCCTGAGACACCTTGG 1111

Qy      365 GGATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCCTGAAGGAGAACGAGACGCGTGTTCGC 424
             |  |||    ||||||||                  ||   ||| ||  ||||
Db     1112 AGACGAGGAGGACCAAGCCATGCGTGCCTTTGAGCTAATGAGGAGCAACGCGGCCCTGTTCCA 1171

Qy      425 GCTGTGCTTCGTGCCCTTCGTGTCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGA 484
             |||| || ||  ||| ||||   |||     || | || |||||||||||||   | |
Db     1172 GCTGGGCTCGGCCCCCCCGGGTGTGTGGATCGTGCTGTGCACGACTCTGAAGCTGCAGATGGA 1231

Qy      485 GCTCGGTCGGGAGCACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCAT 544
             ||||  ||   |||||||   | ||     ||||   ||    ||||   ||||| ||  || |
Db     1232 GAAGGGGGAGAGACCCGGTCCCCACCTGCCTCACCCGCACGGGGCTGTTCCTGCGGTTCCT 1291

Qy      545 CACCAGCGTTCTGAGCTCGGTCCGGTAGCCGACGGGCCCCCGGTTGCAGGGCGACCTGCG 604
             | |||                    |  |||   ||| || || |||||||||   |||
Db     1292 CTGCAGC----------CGGTTCCCGCAGGGCGCACAGCTGCGGGGGCGGCCGCCTGCG 1336

Qy      605 CAATCTGTGCGCCCTGGCCCGCGAGGGCGTCCTCGGACCGCAGGGCGCAGTTGCCGAGAA 664
             | |||  |  || || ||   ||||||||||  | | |||  |   |||
Db     1337 GACGCTGAGCCTTCCTCGGCCCGGCGCAGAGGGCGTGGGCGCAGGTCGTCCGTGCTTCACCGAGA 1396

Qy      665 GGAACTGGAGGCAACTGGAGCTTCGTGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAA 724
             ||||  ||||   |||||  |||      ||||    || | ||||
Db     1397 GGATCTGGAAAGGCTCGGGGTGCAGAGAGTCCGACCTCC---GTCTGTCCTGGACGGGAGA 1453

Qy      725 GGAGCTGCCGGGGCGGGCGTGCTGGAGACAGAGGTCACTCACGTTCATCGACCAGAGCTTCCA 784
             ||||||||  | |  |||||  || |||||||||||   | ||| |  |||||||||||| ||
Db     1454 CATCCTCCGCGACGACAGAGTCTCCAAAGGCTGCTACTCCTTCATCCACCTGCCTTCAGCTTCCA 1513

Qy      785 GGAGTTCCTCCGCGGCACTGTTCTACCTGCTGGAGGACGGCGGGGG 828
             | ||||  || ||  || |||| |||   |||  |||||||||
Db     1514 GCAGTTTTCTCACTGCCCTGTTCTACACCCCTGGAGAAGGAGGAGG 1557
```

Search completed: October  3, 2003, 13:42:26
Job time : 649 secs

GenCore version 5.1.6

Copyright (c) 1993 - 2003 Compugen Ltd.

GX nucleic - nucleic search, using sw model

Run on:    October  3, 2003, 03:53:26 ; Search time 156 seconds
                                          (without alignments)
                                          7606.200 Million cell updates/sec

Title:          US-10-029-347-1
Perfect score:  2689
Sequence:       1 cggaccgcgtgggcgcgcagc.........aaaaaaaaaaaaaaaaaa 2689

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       569978 seqs, 220691566 residues

Total number of hits satisfying chosen parameters:    1139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :      Issued Patents NA:*
                1: /cgn2_6/prodata/2/ina/5A_COMB.seq:*
                2: /cgn2_6/prodata/2/ina/5B_COMB.seq:*
                3: /cgn2_6/prodata/2/ina/6A_COMB.seq:*
                4: /cgn2_6/prodata/2/ina/6B_COMB.seq:*
                5: /cgn2_6/prodata/2/ina/PCTUS_COMB.seq:*
                6: /cgn2_6/prodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|---|
| 1 | 427.6 | 15.9 | 678 | | 4 | US-09-016-434-1011 | Sequence 1011, Ap |
| 2 | 236.6 | 8.8 | 257 | | 4 | US-09-016-434-208 | Sequence 208, App |
| 3 | 85.4 | 3.2 | 1371 | | 2 | US-08-910-731-1 | Sequence 1, Appli |
| 4 | 85.4 | 3.2 | 1371 | | 2 | US-08-910-731-7 | Sequence 7, Appli |
| 5 | 85.4 | 3.2 | 1371 | | 2 | US-08-795-395-1 | Sequence 1, Appli |
| 6 | 69.8 | 2.6 | 1386 | | 4 | US-08-910-731-5 | Sequence 5, Appli |
| 7 | 61.2 | 2.3 | 1470 | | 4 | US-09-099-041A-27 | Sequence 27, Appl |
| 8 | 61.2 | 2.3 | 1470 | | 4 | US-09-245-281-27 | Sequence 27, Appl |
| 9 | 61.2 | 2.3 | 1470 | | 4 | US-09-207-359B-27 | Sequence 27, Appl |
| 10 | 61.2 | 2.3 | 1470 | | 4 | US-09-340-620A-27 | Sequence 27, Appl |
| 11 | 61.2 | 2.3 | 2859 | | 4 | US-09-099-041A-9 | Sequence 9, Appli |
| 12 | 61.2 | 2.3 | 2859 | | 4 | US-09-245-281-9 | Sequence 9, Appli |
| 13 | 61.2 | 2.3 | 2859 | | 4 | US-09-207-359B-9 | Sequence 9, Appli |
| 14 | 61.2 | 2.3 | 2859 | | 4 | US-09-340-620A-9 | Sequence 9, Appli |
| 15 | 61.2 | 2.3 | 3080 | | 4 | US-09-099-041A-25 | Sequence 25, Appl |
| 16 | 61.2 | 2.3 | 3080 | | 4 | US-09-245-281-25 | Sequence 25, Appl |
| 17 | 61.2 | 2.3 | 3080 | | 4 | US-09-207-359B-25 | Sequence 25, Appl |
| 18 | 61.2 | 2.3 | 3080 | | 4 | US-09-340-620A-25 | Sequence 25, Appl |
| 19 | 61.2 | 2.3 | 3382 | | 4 | US-09-245-281-7 | Sequence 7, Appli |
| 20 | 61.2 | 2.3 | 3382 | | 4 | US-09-099-041A-7 | Sequence 7, Appli |
| 21 | 61.2 | 2.3 | 3382 | | 4 | US-09-207-359B-7 | Sequence 7, Appli |
| 22 | 61.2 | 2.3 | 3382 | | 4 | US-09-340-620A-7 | Sequence 7, Appli |
| 23 | 61.2 | 2.3 | 4302 | | 4 | US-09-245-281-38 | Sequence 38, Appl |
| 24 | 61.2 | 2.3 | 4302 | | 4 | US-09-207-359B-38 | Sequence 38, Appl |
| 25 | 61.2 | 2.3 | 4302 | | 4 | US-09-340-620A-38 | Sequence 38, Appl |
| 26 | 61.2 | 2.3 | 32042 | | 4 | US-09-245-281-44 | Sequence 44, Appl |
| 27 | 61.2 | 2.3 | 32042 | | 4 | US-09-340-620A-63 | Sequence 63, Appl |

| 28 | 56.6 | 2.1 | 1374 | | 2 | US-08-910-731-3 | Sequence 3, Appli |
| 29 | 56.6 | 2.1 | 1374 | | 2 | US-08-795-395-3 | Sequence 3, Appli |
| 30 | 55.2 | 2.1 | 1926 | | 4 | US-09-249-585A-2 | Sequence 2, Appli |
| 31 | 55.2 | 2.1 | 1926 | | 4 | US-09-410-399-3 | Sequence 3, Appli |
| 32 | 55.2 | 2.1 | 2580 | | 3 | US-09-050-863-2 | Sequence 2, Appli |
| 33 | 55.2 | 2.1 | 2580 | | 4 | US-09-359-081-2 | Sequence 2, Appli |
| 34 | 55.2 | 2.1 | 5452 | c | 4 | US-09-130-114-1 | Sequence 1, Appli |
| 35 | 55.2 | 2.1 | 8705 | c | 4 | US-09-647-344A-14 | Sequence 14, Appl |
| 36 | 55.2 | 2.1 | 9600 | | 3 | US-08-910-647-1 | Sequence 1, Appli |
| 37 | 55.2 | 2.1 | 9600 | | 4 | US-09-620-925-1 | Sequence 1, Appl |
| 38 | 55.2 | 2.1 | 10596 | | 1 | US-07-884-811-15 | Sequence 15, Appl |
| 39 | 55.2 | 2.1 | 10596 | | 1 | US-07-885-97-15 | Sequence 15, Appl |
| 40 | 55.2 | 2.1 | 10596 | | 1 | US-08-087-783A-15 | Sequence 15, Appl |
| 41 | 55.2 | 2.1 | 10596 | | 1 | US-08-194-088B-15 | Sequence 15, App |
| 42 | 55.2 | 2.1 | 10596 | | 1 | US-08-194-087-15 | Sequence 15, Appl |
| 43 | 55.2 | 2.1 | 690 | | 5 | PCT-US93-04648-15 | Sequence 15, Appl |
| 44 | 53.8 | 2.0 | 1905 | | 4 | US-09-252-991A-15944 | Sequence 15944, A |
| 45 | 53.8 | 2.0 | 1905 | c | 4 | US-09-252-991A-16561 | Sequence 16561, A |

ALIGNMENTS

RESULT 1
US-09-016-434-1011

; Sequence 1011, Application US/C9016434
;  Patent No. 6500938
;; GENERAL INFORMATION:
;  APPLICANT: Janice Au-Young
;  APPLICANT: Jeffrey J. Seilhamer
;  TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
;  TITLE OF INVENTION: PATHWAY GENE EXPRESSION
;  NUMBER OF SEQUENCES: 1490
;  CORRESPONDENCE ADDRESS:
;     ADDRESSE: INCYTE PHARMACEUTICALS, INC.
;     STREET: 3174 PORTER DRIVE
;     CITY: PALO ALTO
;     STATE: CALIFORNIA
;     COUNTRY: USA
;     ZIP: 94304
;  COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;  CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/09/016,434
;     FILING DATE: HEREWITH
;     CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;     APPLICATION NUMBER:
;     FILING DATE:
;     CLASSIFICATION:
;  ATTORNEY/AGENT INFORMATION:
;     NAME: Zeller, Karen J.
;     REGISTRATION NUMBER: 37,071
;     REFERENCE/DOCKET NUMBER: PA-0002 US
;  TELECOMMUNICATION INFORMATION:
;     TELEPHONE: (650) 855-0555
;     TELEFAX: (650) 845-4166
;  INFORMATION FOR SEQ ID NO: 1011:
;     SEQUENCE CHARACTERISTICS:
;        LENGTH: 678 base pairs
;        TYPE: nucleic acid
;        STRANDEDNESS: single
;        TOPOLOGY: linear
;     IMMEDIATE SOURCE:
;        LIBRARY: BRAINOT04
;        CLONE: 927003
US-09-016-434-1011

  Query Match             15.9%;   Score 427.6;  DB 4;  Length 678;
  Best Local Similarity   89.3%;   Pred. No. 1.7e-73;

Matches 550; Conservative 0; Mismatches 57; Indels 9; Gaps 0;

```
Qy   412  AGACGCTGTCGCGCTGCTCGTGTGCTGATCGTGTGCACCGTGCTGC  471
          ||||||  |||| ||  ||||||| | || || |||| |||
Db    53  ANACCCATANCNCNCTGTTCTTCTGTCCCTTCGTGTGGATCGTGTGCACCGTNCTGC  112

Qy   472  GCCAGCAGCTGGAGCTCGGTCGGGACCTGGTCCAA-GACCACCACGTCAGTG  530
          | |||| ||||| ||||||||||||| |||||| |||||||||||| ||||
Db   113  GCCAGCAACTGGAGCTCGGTCGGCGGACCTGTCCGCGACCTCCAAGACCACCACGTCAGTG  172

Qy   531  TACCTGCTTTCATCACCAGCGTTCTGAGCTCGGTTAGCCGACGGGCCCCGGTTG  590
          |||| |||  ||||||||||  ||   ||||||||| || || |||| ||||
Db   173  TACCTGCTTTCATCACCANNTTCTGAGCTCGGTCCGGTAGCCGACGGGCCCCGGTTG  232

Qy   591  CAGGGCGACCTGCGCAATCTGTGCCGCCTGGCCCGGAGGGGTCTTCGGACGCAGGGCG  650
          |||| ||||| ||||||| | |  | || || ||||| | |||||| |||||||||
Db   233  CAGGGCGACCTGCGCAATCTGTGCCGCCTGGCCCGGAGGGGCCTCCTGGACGCAGGGCG  292

Qy   651  CAGTTTGCCGAGAAGGAACTGGAGCAACTGGAGCTTCGTCGGCCCAAAGTGCAGACCGTG  710
          ||  |||||| || ||||| | | |||| |||| ||||||||| |||||||| |||| |
Db   293  CANTTTGCCGAGAAGGAACTGGAGCAACTGGAGCTTCGTGCGGCCCAAAGTGCAGACCGTG  352

Qy   711  TTTCTCAGCAAAAAGGAGCTGCCCGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATC  770
          ||||| || |||||| ||  | | |||||||||| ||  ||| |||| ||||||||| |
Db   353  TTTCTCANCAAAAAGGA-CTGCCCGGGCGTGCTGGAAACAGAGGTCACCTACCAGTTCATC  411

Qy   771  -GACCAGAGCTTCCAGGAGTTCCT-CGCGGCCACTGTCCTACCTGCTGAGGACGGCGGGG  828
           |||||||||| |||| ||| |  | |||||| || | |||||||||||| |||||
Db   412  GGACCAGAGCTTCNAGGAATTCCTCCGCGGCCACTGTCCTACCTGCTGAAGACGGNGGGG  471

Qy   829  TGCCCAGGACCGGCGCTGGCGGCG-TTGGGACACTCCCGCGTGGGACGCCGCCAGCCGCAC  887
          |||| |||| ||| |||| || || ||||| ||||| ||||||| ||  |||
Db   472  TGCCCATGACCGGCGCTGGCGGCGTTGGGACACTCCGCTCCGGGACGCCA-CCGCAC  530

Qy   888  AGCCACTTGGTGTCTCACCACGCGCTTCCTCTTCGGACTTGCTGAGCGCGGAGCCGATGCGC  947
          |||||||||||  || |||| | | | |||||| |||| ||||||||  | ||||| |
Db   531  AGCCACTTGGTGTCCA-CACCCGCTTCCTCTTCCGACTTGCTGAAAGNTGAGC-GAAGCGN  588

Qy   948  GACATCGAGCGCCACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGCAGGACCCTCGG  1007
          |||| || || |||  || ||| |||| |||||| | |||  || ||  ||||||
Db   589  NACATCNACCCCCA-TTCCGCTGCTCATGTTTCCAACNTTTTAACCAGAAGCCCTGCGGT  647

Qy  1008  TGGGTTGCAGGGACAGG  1023
          ||||| ||||| |||||
Db   648  GGGGTTTCAGGGAACAGG  663
```

RESULT 2
US-09-016-434-208
; Sequence 208, Application US/09016434
; Patent No. 6500938
; GENERAL INFORMATION:
;  APPLICANT: Janice Au-Young
;  APPLICANT: Jeffrey J. Seilhamer
;  TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
;  TITLE OF INVENTION: PATHWAY GENE EXPRESSION
;  NUMBER OF SEQUENCES: 1490
; CORRESPONDENCE ADDRESS:
;  ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;  STREET: 3174 PORTER DRIVE
;  CITY: PALO ALTO
;  STATE: CALIFORNIA
;  COUNTRY: USA
;  ZIP: 94304
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/09/016,434
;  FILING DATE: HEREWITH
;  CLASSIFICATION:

; PRIOR APPLICATION DATA:
;  APPLICATION NUMBER:
;  FILING DATE:
;  CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
;  NAME: Zeller, Karen J.
;  REGISTRATION NUMBER: 37,071
;  REFERENCE/DOCKET NUMBER: PA-0002 US
; TELECOMMUNICATION INFORMATION:
;  TELEPHONE: (650) 855-0555
;  TELEFAX: (650) 845-4166
; INFORMATION FOR SEQ ID NO: 208:
;  SEQUENCE CHARACTERISTICS:
;   LENGTH: 257 base pairs
;   TYPE: nucleic acid
;   STRANDEDNESS: single
;   TOPOLOGY: linear
;  IMMEDIATE SOURCE:
;   LIBRARY: SINTBST0:
;   CLONE: 1499408
US-09-016-434-208

Query Match          8.8%;  Score 236.6;  DB 4;  Length 257;
Best Local Similarity 96.9%;  Pred. No. 5.8e-37;
Matches 250;  Conservative 0;  Mismatches 7;  Indels 1;  Gaps 1;

```
Qy   344  GTATTCTACAAGTTCTTCCGGGATGAGAGAGGAGGGCCGAGGCGCGCTACCGCTTCGTGAA  403
          ||||| ||||||||||||||||| ||||||||||||||||||| |||||||||||||||||
Db     1  GTATTTCACAAGTATTTCCGGGATGAGAGAGGAGGGCCGAGGCGCGCTACCGCTTCGTGAA  60

Qy   404  GGAGAACGAGAGACGCTGTTCGCGCTGCGTGTGCCCTTCGTTCGTGCGCTGGATCGTGTGCAC  463
          |||||||||||||| ||||||||||||||||||||| |||||||||||||||||||||||||
Db    61  GGAGAACGAGAGACGCTGTTCGCGCTGCGTGTGCCCTTCGTTCGTGCGCTGGATCGTGTGCAC  120

Qy   464  CGTGCTGCGCCAGCAGCTGAGCTCGGTCGGCGCACGTCAAGACGTCAAGACCACCAC  523
          |||||||||||| |  | |||||||||||||  ||||||||||| ||||| ||||| |
Db   121  CGTGCTGCGCCA-NANTGNGAGCTCGGTCGGGACGTCGGCGCACGTCCGCCAAGACCACCAC  179

Qy   524  GTCAGTGTACCCGCTTTTCATCACCAGCGTTCTGAGCTCGGTAGCCGACGGGCC  583
          ||||||||||||||||||||||||||||||||||||| ||||||||||||||||
Db   180  GTCAGTGTACCTGCTTTTCATCACCAGCGTTCTGAGCTCGGTAGCCGACGGGCC  239

Qy   584  CCGGTTGCAGGGCGACCT  601
          ||||||||||||||||||
Db   240  CCGGTTGCAGGGCGACCT  257
```

RESULT 3
US-08-910-731-1
; Sequence 1, Application US/08910731:
; Patent No. 5932440
; GENERAL INFORMATION:
;  APPLICANT: CHATTERJEE, DEB K.
;  APPLICANT: SHANDILYA, HARINI
;  TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
;  NUMBER OF SEQUENCES: 16
; CORRESPONDENCE ADDRESS:
;  ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;  STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;  CITY: WASHINGTON
;  STATE: D.C.
;  COUNTRY: USA
;  ZIP: 20005-3934
; COMPUTER READABLE FORM:
;  MEDIUM TYPE: Floppy disk
;  COMPUTER: IBM PC compatible
;  OPERATING SYSTEM: PC-DOS/MS-DOS
;  SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;  APPLICATION NUMBER: US/08/910,731
;  FILING DATE: (Herewith)
;  CLASSIFICATION: 435
; PRIOR APPLICATION DATA:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,731
FILING DATE: (Herewith)
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/795,395
FILING DATE: 04-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/794,546
FILING DATE: 03-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,057
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.3440003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1371 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
US-08-910-731-7

Query Match               3.2%;   Score 85.4;  DB 2;  Length 1371;
Best Local Similarity    54.7%;   Pred. No. 7.7e-08;
Matches 192; Conservative   0; Mismatches 156; Indels   3; Gaps   1:

Qy   1471 TCACGCTGTCCACTGCAAACTCCCTGACGCGG'CTGCCGAGACCTTTCTGAGGCGCCTGA 1530
          ||||| |||||||||||||| |||| |||||||||||||||||| ||| |||||||||||
Db     83 TCAGGCTCGACGACTGCGGCCTCACGGAGGCACTGCAAGGACATCGGTTCTGCCCCTCC 142

Qy   1531 GGGCAGCCCCGCACTGACGGAGCTGGGCCTCCTCCACCAACAGGCTGGGC 1590
          |||||||  ||| | |||||||||||||||||||
Db    143 GGGCCAACCCCTCCCTGACGAGC'CTGCTCCGCACCAACAGCTGGGCGATGCCGGCG 202

Qy   1591 TGCGTATGCTGAGTGAGGGCGGCTCCAGTAGCTTCGGCAGAGCC 1650
          |||||  |||||||||| ||||||||||||
Db    203 TGCACCTGGTGCTGCAGGGCTGCAGGCCCCACCTGCTACCTGGGGT 262

Qy   1651 AGCTGCCTGACC----CCCAGCGAGGCCGGAGGCGGGCTGCCGGTATGCTTCGGCGAGAGCC 1707
          ||||||||||||
Db    263 AGAACTGCTCCTGACCACCCTGGATCTCAGCGGCGGGCTGCGCCACTGCGCGCCCC 322

Qy   1708 CCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAATCGCGCCCC 1767
          ||||||||||||||||
Db    323 CCACGCTGCGGGAGCTGCTCCTGGACCCCCAGTGCCACCTGCCACCTGGAGAAGCTGCAGTTGG 382

Qy   1768 TGTGTGCAGTCCTGCAGCACCAGGAGTGCGGCCTGCAGACCCTCAGTCTGG 1818
          ||||||
Db    383 TCTGTGAGGGCTCCTGGACCCCCAGTGCCACCTGGAGAGAGCTGCAGTTGG 433

RESULT 5
US-08-795-395-1
Sequence 1, Application US/08795395
Patent No. 5365399
GENERAL INFORMATION:
APPLICANT: CHATTERJEE, DEB K.
APPLICANT: SHANDILYA, HARINI
TITLE OF INVENTION: Cloning and Expression of Rat Liver and
TITLE OF INVENTION: Porcine Liver Ribonuclease Inhibitor
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

---

APPLICATION NUMBER: 08/795,395
FILING DATE: 04-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/794,546
FILING DATE: 03-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,057
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.3440003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1371 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1...1368
US-08-910-731-1

Query Match               3.2%;   Score 85.4;  DB 2;  Length 1371;
Best Local Similarity    54.7%;   Pred. No. 7.7e-08;
Matches 192; Conservative   0; Mismatches 156; Indels   3; Gaps   1;

Qy   1471 TCACGCTGTCCACTGCAAACTCCCTGACGCGG'CTGCCGAGACCTTTCTGAGGCGCCTGA 1530
          ||||| |||||||||||||| |||| |||||||||||||||||| ||| |||||||||||
Db     83 TCAGGCTCGACGACTGCGGCCTCACGGAGGCACTGCAAGGACATCGGTTCTGCCCCTCC 142

Qy   1531 GGGCAGCCCCGCACTGACGGAGCTGGGCCTCCTCCACCAACAGGCTGGGC 1590
          |||||||  ||| | |||||||||||||||||||
Db    143 GGGCCAACCCCTCCCTGACGAGC'CTGCTCCGCACCAACAGCTGGGCGATGCCGGCG 202

Qy   1591 TGCGTATGCTGAGTGAGGGCGGCTCCAGTAGCTTCGGCAGAGCC 1650
          |||||  |||||||||| ||||||||||||
Db    203 TGCACCTGGTGCTGCAGGGCTGCAGGCCCCACCTGCTACCTGGGGT 262

Qy   1651 AGCTGCCTGACC----CCCAGCGAGGCGGGCTGCCGGTATGGTTCGGCGAGAGCC 1707
          ||||||||||||
Db    263 AGAACTGCTCCTGACCACCCTGGATCTCAGCGGCGGGCTGCGCCACTGCGCGCCCC 322

Qy   1708 CCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAATCGCGCCCC 1767
          ||||||||||||||||
Db    323 CCACGCTGCGGGAGCTGCTCCTGGACCCCCAGTGCCACCTGCCACCTGGAGAAGCTGCAGTTGG 382

Qy   1768 TGTGTGCAGTCCTGCAGCACCAGGAGTGCGGCCTGCAGACCCTCAGTCTGG 1818
          ||||||
Db    383 TCTGTGAGGGCTCCTGGACCCCCAGTGCCACCTGGAGAGAGCTGCAGTTGG 433

RESULT 4
US-08-910-731-7
Sequence 7, Application US/08910731
Patent No. 5932440
GENERAL INFORMATION:
APPLICANT: CHATTERJEE, DEB K.
APPLICANT: SHANDILYA, HARINI
TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:

ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
STREET: 1100 NEW YORK AVE., N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 20005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/795,395
FILING DATE: 04-FEB-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,057
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.3440002
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 1371 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
US-08-795-395-1

Query Match                3.2%;   Score 85.4;  DB 2;  Length 1371;
Best Local Similarity     54.7%;   Pred. No. 7.7e-08;
Matches 192;   Conservative 0;   Mismatches 156;   Indels 3;   Gaps 1;

```
Qy  1471 TCACGGCTGTCCCACTGCAAACTCCCTGACGGCGGTCTGCCGAGACCTTTCTGAGGCCTGA 1530
              ||||| ||||||||||  ||||||||||||   ||||||
Db    93 TCAGGCTCGACGACTGCGGCCTTCACGGAGGAGCACTGCAAGGACATCGGTTCTGCCCTCC 142

Qy  1531 GGGCAGCCCCGGCACTGACCGGGGCTGGCCTTGCCTCCGCCACCAACGAGCTGAGCGTGGC 1590
              ||||||||  ||||||||||| ||||||||   |||||  ||||||| ||||
Db   143 GGGCCAACCCCTCCCTGACCGAGCTCGAGCTGTCGGGCGGTGGGCGATGCCGGGG      202

Qy  1591 TGCGTATGCTGAGTGAGGGCCTAGCCTGGCCTGGGTGCAGGTGCGAGACGGTCAGGGTAC  1650
              ||||  ||||  ||||||   ||  ||||||||    |||||||||||||
Db   203 TGCACGGTGCTGCAGGGCCTGCGCAGGCCCACCTGCAAGATCCAAGATCCCAAGACTCAGCCTGC 262

Qy  1651 AGCTGCCTGACC---CCCAGCGGAGGGGGCTGCGGGGTCCAGGGTATGCTTCGGCAGAGCC 1707
              ||||  ||||     |||||||  |||||||||||  |||||||||||||||||||
Db   263 AGAACTGCTCCCTGACCGAGGCGGGCTGCATCAGCGGCTGGATGTGGGTATGCTTCGGCAGAGCC 322

Qy  1708 CCGCCCTGACCACCCTGGACCTGTCCCAACCTCCCCCACCACCGACCTACC        1767
              |||||||||||||   |||||   ||||  ||||||||||||||||||
Db   323 CCACGCTGCGGGGAGCTGCATCTCAGCGGACGACAACCCACTGGGGGACCGCGGCCTGGCGTGC 382

Qy  1768 TGTGTGAGTCCTGCAGCACCAGGGATGCGGCCTGCGACCCCAGTCTGG         1813
              ||||||||  |||||     ||||   ||   ||||||||||||||
Db   383 TCTGTGAGGGGCTTGACGATGCCAGCTGCCACCTGCCACCCCAGTGTCGCCAGGAAGCTGCAGTTGG 433
```

RESULT 6
US-08-910-731-5
; Sequence 5, Application US/08910731
; Patent No. 5932440
GENERAL INFORMATION:
APPLICANT: CHATTERJEE, DEB K.
APPLICANT: SHANDILYA, HARINI

TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
NUMBER OF SEQUENCES: 16
CORRESPONDENCE ADDRESS:
ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
STREET: 1100 NEW YORK AVE., N.W., SUITE 600
CITY: WASHINGTON
STATE: D.C.
COUNTRY: USA
ZIP: 2005-3934
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/910,731
FILING DATE: (Herewith)
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/795,395
FILING DATE: 04-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/794,546
FILING DATE: 03-FEB-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/024,057
FILING DATE: 16-AUG-1996
ATTORNEY/AGENT INFORMATION:
NAME: ESMOND, ROBERT W.
REGISTRATION NUMBER: 32,893
REFERENCE/DOCKET NUMBER: 0942.3440003
TELECOMMUNICATION INFORMATION:
TELEPHONE: 202-371-2600
TELEFAX: 202-371-2540
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1386 base pairs
TYPE: nucleic acid
STRANDEDNESS: both
TOPOLOGY: both
MOLECULE TYPE: cDNA
US-08-910-731-5

Query Match                2.6%;   Score 69.8;  DB 2;  Length 1386;
Best Local Similarity     51.9%;   Pred. No. 7.5e-05;
Matches 205;   Conservative 0;   Mismatches 187;   Indels 3;   Gaps 2;

```
Qy  1419 CTTCATCCACTCTTCAGGCAATGACTGACCCCACTGTGCCATCTGAGCAGCCTCACGCCTG 1478
              ||| |||| ||   |||   |||   |||||||| ||||| ||  ||| ||||
Db   388 CTGCAGCTGCTCTGCGGAGGACTCCTGACCCCAGTGCCGCCTGGGAAAAGTGCAGCTG     447

Qy  1479 TCCCACTGCGAAACTCCCTGACGGCGGTCTGCGCCAGACCTTTCTGAGGCCCTGTGGGCAGCC 1538
              ||||||||   || |||||||  ||||||||   ||||||||||  ||||||
Db   448 GAGTATTGCAGCCTCTGCCTGGGGATGTGGGACTCAGGGCTGCCAAG             507

Qy  1539 CCCGGACTGACGGAGCTGGGCTTCCTCCACAAGAGGCTCAGTGAGGCGGGACTGCGTATG 1598
              ||||  ||||||||  ||   |||||||||  ||||||||||||||||||
Db   508 CCGGCTCTTCAAGGAGGCTCACGGTTAGCAACATCAATGAGGCGGCGTGTG             567

Qy  1599 CTGAGTGAGGGCAGCAGCCTGGCCCGGCGGGTGCAGCCGTGCAGG--GTACAGCCTGC 1656
              ||||||||||   |||||||||||   |||||||||||   |||   |||||
Db   568 CTGTGCCAGGGCTGAAGGACTCCCCCTGCAAGCTGGGATATGCTTCGGCAGCAAGCTGC 627

Qy  1657 CTGACCCCCAGCGAGGGCT--CCAGTACCTGGTGGTATGCTTCGGCAGAGCCCCCTG 1715
              ||||   ||| |||||    ||||| ||||||  ||||||||||||||||||
Db   628 GGTGTGACATCAGACAACTGCGGGACCTGTGCGGGCATTGTGGCCTCCAAGGGCCTG     687

Qy  1716 ACCACCCTGGATCTGAGCCTGCCAACTGCCGCCCCCATGGTGACCTACCTGTGTGCA 1775
              |||| ||||| ||  |||| ||  |||||||||||||| ||||  ||||||
Db   688 CGGGAGCTGCCCTGGCAGCAACAAGCTGGGTGATGTGGGCATGCGGGAGCTGTGCCCA   747

Qy  1776 GTCCTGCAGCCACCAGGGATGCGGCCTGCCTGCAGACCCT 1810
              |||||| |||||||||| ||||   |||
```

Db        748  GGGCTGCGTCCACCCCAGCTCCAGGCTCAGGACCCCT  782

RESULT 7
US-09-099-041A-27
; Sequence 27, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-076001
; CURRENT FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-099-041A-27

Query Match          2.3%;    Score 61.2;   DB 4;   Length 1470;
Best Local Similarity  50.5%;    Pred. No. 0.0034;
Matches  209;  Conservative   0;  Mismatches 193;  Indels  12;  Gaps   2;

Qy     51  GACCGGCGGCGCCGGTGCCG-AGATGCTGGCCCAGCGGACGGCGGCTGCTTCATCCTG  110
           ||||||||||||||||||| ||||||  ||||||  |||| ||||||||||||||||
Db    673  GACCCCGAGGAGGTGTTGCCTTCCTGCGCT-CCCCAGTGGCCCTTCACCTTC  732

Qy    111  GACGGCGCGGAGGAGCTGCCGGCGCTGGGGGGCCCCGGCCGCGCC---CTGCACAGAC  167
           || |||| |||||||||||||
Db    733  GATGGCCTGGACGAGCTGCACTCGGACTTGGACCTGAGCGCCGTGCCTGACAGCTCCTGC  792

Qy    168  CCCTTCGAGGCGGCGAGCGGCGCGCGGGTGCTAGGCGGCTGCGAGCGAGGCGCTG  227
           |||| ||| |||  |||||   ||| ||| |||||||||||||||||||
Db    793  CCCTGGGAGCCCTGCCCACCCCCGGTC-TGCTGGCCAACCTGCTCAGTGGGAAGCTGCTC  852

Qy    228  CCCACGGCGCCCTCCTGCTGGTGACCACGCGCGCCCGCACAGGCATCGAGGTCCCGCGCGCAGTTCCTG  912
           || || ||||||| ||||  ||||| ||||||| ||||||||||||
Db    853  AAGGGGGCTAGCAAGCTGCTCACAGCCCGCACAGGCATCGAGGTCCCGGCCAGTTCCTG  912

Qy    288  CTGTGTTCCCGCAGTGCGCCGAGGTGCGCGCCTTCTCCGACAAGGACAAGAACAAGAAGTAT  347
           |||||||||||||||||                 ||| || |||||| ||| |||
Db    913  CGGA-------AGAAGGTGCTTCTCCGGGCTTCTCCCCCAGCCACCTGCGCGCCTAT  963

Qy    348  TTCTACAAGTTCTTCCGGGATGAGAGAGGAGGGCCGCCGCCTACCGCTTCGTGAAGGAG  407
           |||| ||||||||||||||  ||| ||||||
Db    964  GCCAGGAGGATGTTCCCCGAGGAGGGCCCTGCAGACGACCGCCTGCTGAGCCAGCTGGAGGCC  1023

Qy    408  AACGAGACGCTGTTCGGCGTGTTGCTTCGGTGCTTGCTGTGGTGCTGGATCGTGTGC  461
           || || |||||  |||||
Db   1024  AACCCCAACCTGCAGCCTGTGCTCTGTGCCCCTCTTCTGTGGATCATCTTC  1077


RESULT 8
US-09-245-281-27
; Sequence 27, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942

---

; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-245-281-27

Query Match          2.3%;    Score 61.2;   DB 4;   Length 1470;
Best Local Similarity  50.5%;    Pred. No. 0.0034;
Matches  209;  Conservative   0;  Mismatches 193;  Indels  12;  Gaps   2;

Qy     51  GACCGGCGGCGCCGGTGCCGCAGATGCTGGCCCAGCGGACGGCGGCTGCTTCATCCTG  110
           ||||||||||||||||||  |||||  ||||||  |||| ||||||||||||||
Db    673  GACCCCGAGGAGGTGTTGCCTTCCTGCGCT-CCCCAGTGGCCCTTCACCTTC  732

Qy    111  GACGGCGCGGAGGAGCTGCCGGCGCTGGGGGGCCCCGGCCGCGCC---CTGCACAGAC  167
           || |||| |||||||||||||
Db    733  GATGGCCTGGACGAGCTGCACTCGGACTTGGACCTGAGCGCCGTGCCTGACAGCTCCTGC  792

Qy    168  CCCTTCGAGGCGGCGAGCGGCGCGCGGGTGCTGAGCGGCTGCGAGTAAGGCGCTG  227
           |||| ||| |||
Db    793  CCCTGGGAGCCCTGCCCACCCCCGGTC-TGCTGGCCAACCTGCTCAGTGGGAAGCTGCTC  852

Qy    228  CCCACGGCGCCCTCCTGCTGGTGACCACGCGCGCCCGCACAGGCATCGAGGTCCCGCGCGCAGTTCCTG  912
           || || ||||||| |||| ||||| ||||||||
Db    853  AAGGGGGCTAGCAAGCTGCTCACAGCCCGCACAGGCATCGAGGTCCCGGCCAGTTCCTG  912

Qy    288  CTGTGTTCCCGCAGTGCGCCGAGGTGCGCGCCTTCTCCGACAAGGACAAGAACAAGAAGTAT  347
           ||||||||||
Db    913  CGGA-------AGAAGGTGCTTCTCCGGGCTTCTCCCCCAGCCACCTGCGCGCCTAT  963

Qy    348  TTCTACAAGTTCTTCCGGGATGAGAGAGGAGGGCCGCCGCCTACCGCTTCGTGAAGGAG  407
           |||| ||
Db    964  GCCAGGAGGATGTTCCCCGAGGAGGGCCCTGCAGACGACCGCCTGCTGAGCCAGCTGGAGGCC  1023

Qy    408  AACGAGACGCTGTTCGGCGTGTTGCTTCGGTGCTTGCTGTGGTGCTGGATCGTGTGC  461
           || ||
Db   1024  AACCCCAACCTGCAGCCTGTGCTCTGTGCCCCTCTTCTGTGGATCATCTTC  1077


RESULT 9
US-09-207-359B-27
; Sequence 27, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 27
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-207-359B-27

Query Match          2.3%;    Score 61.2;   DB 4;   Length 1470;
Best Local Similarity  50.5%;    Pred. No. 0.0034;
Matches  209;  Conservative   0;  Mismatches 193;  Indels  12;  Gaps   2;

Qy     51  GACCGGCGGCGCCGGTGCCGCAGATGCTGGCCCAGCGGACGGCGGCTGCTTCATCCTG  110
           ||||||||||||||||||  |||||  ||||||  |||| ||||||||||||||
Db    673  GACCCCGAGGAGGTGTTGCCTTCCTGCGCT-CCCCAGTGGCCCTTCACCTTC  732

## Left column

```
Qy  111 GACGGCGCGGACGAGCTGCCCGGGCTGGGGGGCCCCGAGGCCCGGCC---CTGCACAGAC 167
Db  733 GATGGCCTGGACGAGCTGCACTCGGACTTGGACCTCGAGCCCGGTCGGTGCTGACAGCTCCTGC 792
Qy  168 CCCTTCGGAGGGCCCTGTGCGCGCGGGTGCTGGGGGTGCTGAGTAAGGCGCTGCTG 227
Db  793 CCCTGGGAGCCTGCCACCCCTGGTCTTGCTGCCAACCTGCTGCCCGGGGAGGCTGGGAAGCTCCTC 852
Qy  228 CCCACGGCCTCCTGCGTGGTGACCACCCCTGCTGCCGGGCGCGCGCACAGGCATCGAGGTCCCGGCCCAGTCCTG 287
Db  853 AAGGGGGCCTAGCAAGCTGCTCACAGCCCGCACCAGAGGTCCCGGCCCAGTCCTG 912
Qy  288 CCGTGTTCCCCGCAGTGCCGCGAGGTGCGCGCCTTCCCGACAAGAAGACAAGAAGAGTAT 347
Db  913 CGGA--------AGAAGGTGCTTCCCGGGGCTTCTCCCCCAGCCACCTCGCGGCGCCTA 963
Qy  348 TTCTACAAGTTCTCCCGGGATGAGAGAGGAGGGGGCCCGGCCTACCGCTTCGTGAAGGAG 407
Db  964 GCCAGGAGGATGTTCCCCGACGCGGGGCCCGCAGGACCGCTTCGTGTCTGTGATCGTGTGC 1023
Qy  408 AACGAGACGCGTGTTCGGCGTGCTTCGTGTGTCTGATCGTGTGC 461
Db 1024 AACCCCAACCTCTGCAGCCTGTGCTCCTGTGCTGTGGATCATCTTC 1077
```

RESULT 10
US-09-340-620A-27
; Sequence 27, Application US/09340620A
; Patent No. 6482933
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THERE
; FILE REFERENCE: 07334-124001
; CURRENT APPLICATION NUMBER: US/09/340,620A
; CURRENT FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/245,281
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 7;
; SEQ ID NO 27
; LENGTH: 1470
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-340-620A-27

Query Match         2.3%;  Score 61.2;  DB 4;  Length 1470;
Best Local Similarity  50.5%;  Pred. No. 0.0034;
Matches 209;  Conservative 0;  Mismatches 193;  Indels 12;  Gaps 2;

```
Qy   51 GACCGGCGGCGCCGGTGCCGCAGATGCTGGCCCAGCGGCCAGCCGGCGCCGAGCGGCTGCTGCTTCATCCTG 110
Db  673 GACCCCGAGGAGGTGTTTGCCTTCCTGCTGCCCTTTGGCGTGCCCCTCTTCCCCACGTGGCCCTCCTTCACCTTC 732
Qy  111 GACGGCGCGGACGAGCTGCCCGGGCTGGGGGGCCCCGAGGCCCGGCC---CTGCACAGAC 167
Db  733 GATGGCCTGGACGAGCTGCACTCGGACTTGGACCTCGAGCCCGGTCGGTGCTGACAGCTCCTGC 792
Qy  168 CCCTTCGGAGGGCCCTGTGCGCGCGGGTGCTGGGGGTGCTGAGTAAGGCGCTGCTG 227
Db  793 CCCACGGCCCTCCTGCGTGGTGACCACGCGGGCGCGCGCACAGGCATCGAGGTCCCGGCCCAGTCCTG 852
Qy  228 CCCACGGCCTCCTGCGTGGTGACCACCCCTGCTGCCGGGCGCGCGCACAGGCATCGAGGTCCCGGCCCAGTCCTG 287
Db  853 AAGGGGGCCTAGCAAGCTGCTCACAGCCCGCACCAGAGGTCCCGGCCCAGTCCTG 912
Qy  288 CTGTGTTCCCCGCAGTGCCGCGAGGTGCGCGCCTTCTCCGGCGACAAGAAGACAAGAAGAGTAT 347
```

## Right column

```
Db  913 CGGA--------AGAAGGTGCTTCCCGGGGCTTCTCCCCCAGCCACCTCGCGCGCCTAT 953
Qy  348 TTCTACAAGTTCTCCCGGATGAGAGGAGGGGGCCTGAGACCGGCGGCGCCTACCGCTTCGTGAAGGAG 407
Db  964 GCCAGGAGGATGTTCCCGAGCGGAGGGGCCCTGCAGGACCGCTTCGTGTCTGTGATCGTGTGC 1023
Qy  408 AACGAGACGCTGTTCGGCGCTGCTTCGTGTGTCTGATCGTGTGC 461
Db 1024 AACCCCAACCTCTGCAGCCTGTGCTCTGTGGATCATCTTC 1077
```

RESULT 11
US-09-099-041A-9
; Sequence 9, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-076001
; CURRENT APPLICATION NUMBER: US/09/099,041A
; CURRENT FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-099-041A-9

Query Match         2.3%;  Score 61.2;  DB 4;  Length 2859;
Best Local Similarity  50.5%;  Pred. No. 0.0038;
Matches 209;  Conservative 0;  Mismatches 193;  Indels 12;  Gaps 2;

```
Qy   51 GACCGGCGGCGCCGGTGCCGCAGATGCTGGCCCAGCGGCCAGCCGGCGCCGAGCGGCTGCTGCTTCTTCATCCTG 110
Db  790 GACCCCGAGGAGGTGTTTGCCTTCCTGCTGCCCTTTGGCGTGCCCCTCTTCCCCCACGTGGCCCTCCTTCACCTTC 849
Qy  111 GACGGCGCGGACGAGCTGCCCGGGCTGGGGGGCCCCGAGGCCCGGCC---CTGCACAGAC 167
Db  850 GATGGCCTGGACGAGCTGCACTCGGACTTGGACCTCGAGCCCGGTCGGTGCTGACAGCTCCTGC 909
Qy  168 CCCTTCGGAGGGCCCGAGGTGCGCGCGGGTGCTGGGGGTGCTGAGTAAGGCGCTGCTG 227
Db  910 CCCTGGGAGCCTGCCACCCCTGGTCTTGCTGCCAACCTGCTCAGTGGGAGCTGCTCTC 969
Qy  228 CCCACGGCCTCCTGCGTGGTGACCACCCCTGCTGCCGGGCGCGCGCACAGGCATCGAGGTCCCGGCCCAGTCCTG 287
Db  970 AAGGGGGCCTAGCAAGCTGCTCACAGCCCGCACCAGAGGTCCCGGCCCAGTCCTG 1029
Qy  288 CTGTGTTCCCCGCAGTGCCGCGAGGTGCGGATGAGAGGAGGGGGCCTGAGACCGGCGGCGCCTACCGCTTCGTGAAGGAGTAT 347
Db 1030 CGGA--------AGAAGGTGCTTCCCGGGGCTTCTCCCCCAGCCACCTGCGCGCGCCTAT 1080
Qy  348 TTCTACAAGTTCTCCGGATGAGAGGAGGGGGCCTGAGACCGGCGGCGCCTACCGCTTCGTGAAGGAG 407
Db 1081 GCCAGGAGGATGTTCCCGAGCGGAGGGGCCCTGCAGGACCGCTTCGTGTCTGTGATCGTGTGC 1140
Qy  408 AACGAGACGCTGTTCGGCGTGCTTCGTGTGTCTGATCGTGTGC 461
Db 1141 AACCCCAACCTCTGCAGCCTGTGCTCTGTGGATCATCTTC 1194
```

RESULT 12
US-09-245-281-9
; Sequence 9, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF

; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-245-281-9

Query Match       2.3%;   Score 61.2;  DB 4;  Length 2859;
Best Local Similarity   50.5%;   Pred. No. 0.0038;
Matches 209;  Conservative   0;  Mismatches 193;  Indels 12;  Gaps   2;

Qy     51  GACCGGCGGCGGCCGGTGCCGCAGATGCTGGCCCAGCCGCCAGCGGCTGCTCTTCATCCTG  110
               |||| || |  ||||  || ||||  || |||  |||||||||||||||||||||
Db    790  GACCCCGAGGAGGTGTTTGCCTTCGCTTCCCCACGTGGCCCTCTTCACCTTC  849

Qy    111  GACGGCGCGGACGAGCTGCCGGGCGCTGGGGGCCCGAGGCGCGGCC---CTGCACAGAC  167
               ||||||||| |||||  |||| ||  |||||||   | |||||||    |||||||||
Db    850  GATGGCCTGACGACGAGCTGCACTCGGACTTGGACTTGGACCGTGCTGCCTGACAGCTCCTGC  909

Qy    168  CCCTTCGAGGCGGCGAGCGGGCGGGGGCGCTGTGCTGAGTAAGGCGCTGCTG  227
               |||| |||||   | |  |||  ||||  |||||| ||||  |||||||||
Db    910  CCCTGGGAGCCTGCGCCCACCCCCTGGTCTTGCTGGCCAACCTGCTCAGTGGGAAGCTGCTC  969

Qy    228  CCCACGGCCCCTCCTGCTGGTGACCACGCGCCGCCCCCGGGCCGCTGCAGGGCCGC  287
               |||  |||| |||  |  |||||| | | |||||||   ||||  |||||||||
Db    970  AAGGGGGCTAGCAAGCTGCTCACAGCCCCACAGGCATCGAGTCCCGCCCAGTTCCTG  1029

Qy    288  CTGTGTTCCCGCAGTGCGCGAGGTGCGGCGGTTCTCCGACAAGGACAAGAAGTAT  347
               ||| |  |||| |||  || | | || |||||| || ||||| ||||||||||
Db   1030  CGGA--------AGAAGTGCTTCTCCGGGCCTTCTCCCCAGCCACCTGCGGGCCTAT  1080

Qy    348  TTCTACAAGTTCTTCCGGGATGAGGAGGGCGCCTGCAGGAGCCGCCGGAGGAG  407
               |||||||||||||  || ||| | | ||| ||||| |||  ||||||||||
Db   1081  GCCAGGAGGATGTTCCCCGAGCGGCGCCCTGCTGAGCCAGCGGACCAGTGGAGCC  1140

Qy    408  AACGAGACGCTGTTCGGCGCTGTTCGTGCCTTCGTGCCTTCGTGTGTGGATCGTGTGC  461
               ||||| | | ||||  |||  ||  ||  | ||||||| ||||||
Db   1141  AACCCCAACCTGTGCCAGCCTGTTGCTGTGTCTGTCCCTCTTCTGCTGGATCATCTTC  1194

RESULT 14
US-09-340-620A-9
; Sequence 9, Application US/09340620A
; Patent No. 6482933
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THE
; FILE REFERENCE: 07334-124001
; CURRENT APPLICATION NUMBER: US/09/340,620A
; CURRENT FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/245,281
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-340-620A-9

Query Match       2.3%;   Score 61.2;  DB 4;  Length 2859;
Best Local Similarity   50.5%;   Pred. No. 0.0038;
Matches 209;  Conservative   0;  Mismatches 193;  Indels 12;  Gaps   2;

Qy     51  GACCGGCGGCGGCCGGTGCCGCAGATGCTGGCCCAGCCGCCAGCGGCTGCTCTTCATCCTG  110
               |||| || |  ||||  || ||||  || |||  |||||||||||||||||||||
Db    790  GACCCCGAGGAGGTGTTTGCCTTCGCTTCCCCACGTGGCCCTCTTCACCTTC  849

Qy    111  GACGGCGCGGACGAGCTGCCGGGCGCTGGGGGCCCGAGGCGCGGCC---CTGCACAGAC  167
               ||||||||| |||||  |||| ||  |||||||   | |||||||    |||||||||
Db    850  GATGGCCTGACGACGAGCTGCACTCGGACTTGGACTTGGACCGTGCTGCCTGACAGCTCCTGC  909

Qy    168  CCCTTCGAGGCGGCGAGCGGGCGGGGGCGCTGTGCTGAGTAAGGCGCTGCTG  227
               |||| |||||   | |  |||  ||||  |||||| ||||  |||||||||

<hr/>

RESULT 13
US-09-207-359B-9
; Sequence 9, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-207-359B-9

Query Match       2.3%;   Score 61.2;  DB 4;  Length 2859;
Best Local Similarity   50.5%;   Pred. No. 0.0038;
Matches 209;  Conservative   0;  Mismatches 193;  Indels 12;  Gaps   2;

Qy     51  GACCGGCGGCGGCCGGTGCCGCAGATGCTGGCCCAGCCGCCAGCGGCTGCTCTTCATCCTG  110
               |||| || |  ||||  || ||||  || |||  |||||||||||||||||||||
Db    790  GACCCCGAGGAGGTGTTTGCCTTCGCTTCCCCACGTGGCCCTCTTCACCTTC  849

Qy    111  GACGGCGCGGACGAGCTGCCGGGCGCTGGGGGCCCGAGGCGCGGCC---CTGCACAGAC  167
               ||||||||| |||||  |||| ||  |||||||   | |||||||    |||||||||
Db    850  GATGGCCTGACGACGAGCTGCACTCGGACTTGGACTTGGACCGTGCTGCCTGACAGCTCCTGC  909

Qy    168  CCCTTCGAGGCGGCGAGCGGGCGGGGGCGCTGTGCTGAGTAAGGCGCTGCTG  227
               |||| |||||   | |  |||  ||||  |||||| ||||  |||||||||
Db    910  CCCTGGGAGCCTGCGCCCACCCCCTGGTCTTGCTGGCCAACCTGCTCAGTGGGAAGCTGCTC  969

Qy    228  CCCACGGCCCCTCCTGCTGGTGACCACGCGCCGCCCCCGGGCCGCTGCAGGGCCGC  287
               |||  |||| |||  |  |||||| | | |||||||   ||||  |||||||||
Db    970  AAGGGGGCTAGCAAGCTGCTCACAGCCCCACAGGCATCGAGTCCCGCCCAGTTCCTG  1029

Qy    288  CTGTGTTCCCGCAGTGCGCGAGGTGCGGCGGTTCTCCGACAAGGACAAGAAGTAT  347
               ||| |  |||| |||  || | | || |||||| || ||||| ||||||||||
Db   1030  CGGA--------AGAAGTGCTTCTCCGGGCCTTCTCCCCAGCCACCTGCGGGCCTAT  1080

Qy    348  TTCTACAAGTTCTTCCGGGATGAGGAGGGCGCCTGCAGGAGCCGCCGGAGGAG  407
               |||||||||||||  || ||| | | ||| ||||| |||  ||||||||||
Db   1081  GCCAGGAGGATGTTCCCCGAGCGGCGCCCTGCTGAGCCAGCGGACCAGTGGAGCC  1140

Qy    408  AACGAGACGCTGTTCGGCGCTGTTCGTGCCTTCGTGCCTTCGTGTGTGGATCGTGTGC  461
               ||||| | | ||||  |||  ||  ||  | ||||||| ||||||
Db   1141  AACCCCAACCTGTGCCAGCCTGTTGCTGTGTCTGTCCCTCTTCTGCTGGATCATCTTC  1194

; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-207-359B-9

```
Db    910  CCCTGGGAGCCTGCCCACCCCTGGTCTTGCTGGCCAACCTGCTCAGTGGGAGCTGCTC  959

Qy    228  CCCACGGCGCCTCCTGCTGGTGACCACGGCCCGGCGCCCCGGGAGGCTGCAGGGCCGC  287

Db    970  AAGGGGGCTAGCAAGTGCTCACAGCCCGACAGGCATCGAGGTCCGGCGCCAGTTCCTG  1029

Qy    288  CTGTGTTCCCCGCAGTGCGCGAGGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGTAT  347

Db   1030  CGGA------AGAAGGTGCTTCTCCGGGGCTTCTCCCCAGCCACCTGCGCGCCCAT  1080

Qy    348  TTCTACAAGTTCTTCGGGATGAGAGAGGGCCGAGCGCGCCCTACCGCTTCGTGAAGGAG  407

Db   1082  GCCAGGAGGATGTTCCCCGACGGGGCCCTGCAGGACCGCCTGCTGAGCCAGCTGGAGGCC  1140

Qy    408  AACGAGACGCGTGTTCGCGCTGTCGTTCGTGCCCTTCGTGTGCTGATCGTGTGC  461

Db   1141  AACCCCAACCTCTGCAGCCTGGTCTCTGTGCCCCTCTTCTGCTGGATCATCTTC  1194
```

RESULT 15
US-09-099-041A-25
; Sequence 25, Application US/0909904LA
; Patent No. 6340576
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;   TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;   FILE REFERENCE: 07334-076001
;   CURRENT APPLICATION NUMBER: US/09/099,041A
;   CURRENT FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 09/019,942
;   PRIOR FILING DATE: 1998-02-06
;   NUMBER OF SEQ ID NOS: 37
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 25
;   LENGTH: 3080
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (1)...(1470)
US-09-099-041A-25

Query Match       2.3%;   Score 61.2;  DB 4;    Length 3080;
Best Local Similarity   50.5%;    Pred. No. 0.0039;
Matches  209;   Conservative   0;  Mismatches 193;   Indels  12;   Gaps   2;

```
Qy     51  GACCGGCGGCGGGCCGGCGATCCTGGCCCGAGATCCTGGCCCAGCCGGCGGCTGCTTCATCCCG  110
Db    673  GACCCCGAGGAGGAGTGTTTGCCTTCCCGCTGCGGCTTCCCCCACGTGCCCCTTCTTCACCTTC  732

Qy    111  GACGGGCCGGACGAGAGCTGCCCGGCGCTGGGGGGCGCCCGGCCGGCC--CTGCACAGAC  167
Db    733  GATGGCCTGGACGAGCTGCACTCGGACTTGGACTTGGACCTGGACTTGGACCTGCTGACAGCTCCTGC  792

Qy    168  CCCTTGGAGCGGCGGCGGGCGGAGCGGCGGCGGCGGGTGCTGAGCTAGGCGGTGCTGCTG  227
Db    793  CCCTGGGAGCCGAGCGGGTGCGCCCCACCCCCTGGTCTTGGCCAACCTGCTCAGTGGGAAGCTGCTC  852

Qy    228  CCCACGGCGCCTCCTGCTGGTGACCACGGGCCGGGGGGAGGCTGCAGGGCCGC  287
Db    853  AAGGGGGCTAGCAAGTGCTCACAGCCCGACAGGCATCGAGGTCCGGCGCCAGTTCCTG  912

Qy    288  CTGTGTTCCCCGCAGTGCGCGAGGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGTAT  347
Db    913  CGGA------AGAAGGTGCTTCTCCGGGGCTTCTCCCCAGCCACCTGCGCGCCCAT  963

Qy    348  TTCTACAAGTTCTTCGGGATGAGAGAGGGCCGAGCGCGCCCTACCGCTTCGTGAAGGAG  407
Db    964  GCCAGGAGGATGTTCCCCGACGGGGCCCTGCAGGACCGCCTGCTGAGCCAGCTGGAGGCC  1023

Qy    408  AACGAGACGCGTGTCGCGCCTTCGTGTGCTGATCGTGTGC  461
```

Db   1024  AACCCCAACCTCTGCAGCCTGGTCTCTGTGCCCCTCTTCTGCTGGATCATCTTC  1077

Search completed: October  3, 2003, 07:00:16
Job time : 158 secs

OM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on:         October  3, 2003, 06:45:55 ; Search time 5518 Seconds
                (without alignments)
                11843.917 Million cell updates/sec

Title:          US-10-029-347-1
Perfect score:  2689
Sequence:       1 cggacgcgtgggcgcgcagc.........aaaaaaaaaaaaaaaaaaaaa 2689

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       22781392 seqs, 12152238056 residues

Total number of hits satisfying chosen parameters:    45562784

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    EST:*
       1:    em_estba:*
       2:    em_esthum:*
       3:    em_estin:*
       4:    em_estmu:*
       5:    em_estov:*
       6:    em_estpl:*
       7:    em_estro:*
       8:    em_htc:*
       9:    gb_est1:*
      10:    gb_est2:*
      11:    gb_htc:*
      12:    gb_est3:*
      13:    gb_est4:*
      14:    gb_est5:*
      15:    em_estfun:*
      16:    em_estom:*
      17:    em_gss_hum:*
      18:    em_gss_inv:*
      19:    em_gss_pln:*
      20:    em_gss_vrt:*
      21:    em_gss_fun:*
      22:    em_gss_mam:*
      23:    em_gss_mus:*
      24:    em_gss_pro:*
      25:    em_gss_rod:*
      26:    em_gss_phg:*
      27:    em_gss_vrl:*
      28:    gb_gss1:*
      29:    gb_gss2:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 1030.8 | 38.3 | 2730 | 11 | AK039531 | AK039531 Mus musc |
| 2 | 550.8 | 20.5 | 691 | 9 | AI344276 | AI344276 tc03aC7.x |
| C 3 | 411 | 15.3 | 427 | 9 | AI762985 | AI762985 wi05e11.x |
| 4 | 403.8 | 15.0 | 798 | 9 | AW107005 | AW107005 um:8d07.y |

| | | | | | | |
|---|---|---|---|---|---|---|
| 5 | 386 | 14.4 | 719 | 12 | BG971110 | BG971110 602837817 |
| 6 | 339.4 | 12.6 | 603 | 12 | BG971489 | BG971489 602838184 |
| 7 | 337.4 | 12.5 | 938 | 10 | BF783992 | BF783992 602107835 |
| 8 | 316.4 | 11.8 | 670 | 10 | BB633515 | BB633515 BB633515 |
| 9 | 307.6 | 11.4 | 832 | 12 | BG655959 | BG655959 602829753 |
| 10 | 268.8 | 10.0 | 416 | 13 | BY262618 | BY262618 BY262618 |
| C 11 | 201.2 | 7.5 | 394 | 10 | BF561521 | BF561521 UI-R-C0-h |
| 12 | 200.6 | 7.5 | 377 | 9 | AI558213 | AI558213 vj37a07.y |
| 13 | 199.4 | 7.4 | 431 | 13 | BY281505 | BY281505 BY28.505 |
| 14 | 198.8 | 7.4 | 532 | 9 | AA541951 | AA541951 vj37a07.r |
| 15 | 171 | 6.4 | 495 | 13 | BX092006 | BX092006 BX092006 |
| 16 | 167 | 6.2 | 3368 | 11 | BC021272 | BC021272 Homo sapi |
| 17 | 167 | 6.2 | 3369 | 11 | BC012789 | BC012789 Homo sapi |
| 18 | 157.8 | 5.9 | 409 | 9 | AA997492 | AA997492 UI-R-C0-h |
| 19 | 157.4 | 5.9 | 488 | 9 | AI037137 | AI037137 uc90h05.y |
| 20 | 152.8 | 5.7 | 921 | 13 | BQ431649 | BQ431649 AGENCOURT |
| 21 | 138 | 5.1 | 664 | 29 | AG067278 | AG067278 Pan trogl |
| 22 | 132 | 4.9 | 1110 | 13 | BX457482 | BX457482 BX457482 |
| 23 | 130.2 | 4.8 | 1156 | 10 | BE731237 | BE731237 601567129 |
| 24 | 125.6 | 4.7 | 934 | 10 | BE733220 | BE733220 601568455 |
| 25 | 122.8 | 4.6 | 629 | 14 | CB016418 | CB016418 pgnic.pk0 |
| 26 | 122.4 | 4.6 | 939 | 13 | BX371966 | BX371966 BX371966 |
| 27 | 120.8 | 4.5 | 124 | 14 | C01661 | C01661 HUMGS000867 |
| 28 | 120.8 | 4.5 | 2218 | 11 | AF054176 | AF054176 Homo sapi |
| 29 | 113 | 4.2 | 440 | 28 | AZ940081 | AZ940081 2X0199013 |
| C 30 | 101.2 | 3.8 | 577 | 28 | AZ432279 | AZ432279 1M0217D08 |
| 31 | 99.2 | 3.7 | 646 | 10 | BE265950 | BE265950 601193329 |
| 32 | 96.6 | 3.6 | 2394 | 11 | AK028575 | AK028575 Mus muscu |
| 33 | 96.6 | 3.6 | 2520 | 11 | AK032446 | AK032446 Mus muscu |
| 34 | 96.6 | 3.6 | 4476 | 11 | AK028802 | AK028802 Mus muscu |
| C 35 | 92 | 3.4 | 491 | 9 | AI925881 | AI925881 wt:3a09.x |
| 36 | 89 | 3.3 | 506 | 14 | CA556855 | CA556855 K0218D11- |
| 37 | 88.6 | 3.3 | 587 | 9 | AV741792 | AV741792 AV741792 |
| 38 | 86.2 | 3.2 | 3275 | 11 | AK054426 | AK054426 Mus muscu |
| 39 | 85.4 | 3.2 | 568 | 12 | BI344799 | BI344799 373381 MA |
| 40 | 85.4 | 3.2 | 585 | 12 | BI344410 | BI344410 372877 MA |
| 41 | 84.6 | 3.1 | 589 | 12 | BM030946 | BM030946 495506 MA |
| 42 | 84.6 | 3.1 | 600 | 14 | CA034391 | CA034391 4C0001- B |
| 43 | 83.2 | 3.1 | 595 | 9 | AI883819 | AI883819 fc70d12.y |
| 44 | 83 | 3.1 | 554 | 14 | CB441719 | CB441719 692223 MA |
| 45 | 82.8 | 3.1 | 253 | 14 | CB606610 | CB606610 AMGNNJC:C |

ALIGNMENTS

RESULT 1
AK039531
LOCUS          AK039531          2730 bp    mRNA    linear    HTC 05-DEC-2002
DEFINITION     Mus musculus adult male spinal cord cDNA, RIKEN full-length
               enriched library, clone:A330055K17 product:similar to VASOPRESSIN
               RECEPTOR [Rattus norvegicus], full insert sequence.
ACCESSION      AK039531
VERSION        AK039531.1  GI:26087232
KEYWORDS       HTC; CAP trapper.
SOURCE         Mus musculus (house mouse)
  ORGANISM     Mus musculus
               Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
               Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE      1
  AUTHORS      Carninci,P. and Hayashizaki,Y.
  TITLE        High-efficiency full-length cDNA cloning
  JOURNAL      Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE      99279253
  PUBMED       10349636
REFERENCE      2
  AUTHORS      Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
               Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE        Normalization and subtraction of cap-trapper-selected cDNAs to
               prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL      Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE      20499374
  PUBMED       11042159

```
REFERENCE   3
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,G.,
            Okazaki,Y., Muramatsu,M. Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530913
PUBMED      11076861
REFERENCE   4
AUTHORS     Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Tomita,X.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,X.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bul,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409 (6821), 685-690 (2001)
MEDLINE     21085660
PUBMED      11217851
REFERENCE   5
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6   (bases 1 to 2730)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,T.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,T., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kondo,Y., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shiraagawa,T., Shiraki,T.,
            Sogabe,Y., Tagami,K., Tagawa,T., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,Y.,
            Miramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome
            Exploration Research Group; RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute, 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
```

```
source          1..2730
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="C57BL/6J"
                /db_xref="FANTOM_DB:A330055K17"
                /db_xref="taxon:10090"
                /clone="A330055K17"
                /sex="male"
                /tissue_type="spinal cord"
                /clone_lib="RIKEN full-length enriched mouse cDNA library"
                /dev_stage="adult"
misc_feature    1..2730
                /note="similar to VASOPRESSIN RECEPTOR [Rattus norvegicus]
                (SPTR|Q63035), evidence: FASTY, 82.2%ID, 100%length,
                match=1440)"

BASE COUNT       649 a       730 c       767 g       584 t
ORIGIN

Query Match             38.3%;  Score 1030.8;  DB 11;  Length 2730;
Best Local Similarity   69.6%;  Pred. No. 5.1e-88;
Matches 1536;  Conservative    0;  Mismatches 637;  Indels  33;  Gaps   9;

Qy       7 CGTGGGCGCAGCCTGCTGACCTGACCAGTGCCCGCCGCGGCGCGGCGCGCGCGG  60
           |||||| | ||||||||||||||||| ||| || ||||| ||| ||||| |||||
Db     165 CGGGAAACGGCAGCTTGCGGGACCTGGTGCTGGACCAGTGCCCGGACCGGCTTGGCCAG  224

Qy      67 TGCCGCAGATGCTGGGCCGCCAGCCGAGGGCTGCTTCATCGTGGAGGGGGACGAGC  126
           |||||| | ||||||||| ||||||| || || ||||||||| ||||||| | ||||
Db     225 TGAAGCGGATACTGGGCGCAACCCAATCGCCTACTGTTCATCGTGGACGGTGCGACAGAGC  284

Qy     127 TGCCGGCCTGGGGGGGCCCCGAGGCCGCGGCCCACTCCCTGCACAGACCCCTTCGAGGCCGAGGCG  186
           ||||  || |||||  ||| || ||||||| ||||| ||||  | ||||||||||||| | |||||
Db     285 TGCCTACTTACCATCCTCAGAGCCTCGAGGGGCTGCTGAGTAAGGCGCGTGCTGCCCACGGCGCCTGG  344

Qy     187 GGGCGCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCGTGCTGCCCACGGCGCCCTCCTGCTGG  246
           |||||| |||||||||||||| ||||| |  | |||||||||||  ||| |||| ||||||||||
Db     345 GCTTGAGGTGCTGAGCGGGCTGCTGCGTGAGCCTGAGCCAGGAGACTGTTACCAGGAGCTCGCTTGCTAG  404

Qy     247 TGACCACCAGGCGCGGCGGCGCCCCCGGGCGAGGCGTGCAGGGCTGCAGGCGGTGAGTGTTCCCGCAGTGCG  306
           |||||||||||  |   || || |  |  || |||||| |||  | |   |||||  |||  |  | ||||||
Db     405 TGACTACACGCGCCATGCGCTACAGGGAGGGCTGCAGGGCTGCAGGATTGTGCTCGGCCACAGTGCG  464

Qy     307 CCGAGGTGCGGCGCTTCTCCGACAAGGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG  366
           |||||||||  |||||||| |||||||| ||||||||  |  | ||| |||| ||||||||||||||
Db     465 CAGAAATACGGCGGCTTCTCAGACAAGGACAAGAACAAAAAAAAGTATTTCTTCAAGTTCTTCCGGG  524

Qy     367 ATGAGAGGAGGGCGAGCGGCTACCGCTTCGTGAAGGAGAACGAGACGCGTTCGTCGCGC  425
           |||||| || || |||||||||| |  || || ||| || | || ||| |  ||||||||
Db     525 ACGAGAGGAGGCAGAGCGGCGGCTTACCGCTTCCGTGAAAGAGAACGAGACGCGTTGCAC  584

Qy     427 TGTGCTTCGTGCCCTTCGTGCTGGATCGTCGTGCTGCTCGCCGCCAGCAGCTGGAGC  486
           |||| |||||||||| |||||| |||||||| |||  |||| |||| ||| ||| |||
Db     585 TGTGCTTTGTGCCCTCCGTGTGCTGGATCGTCGTGCTGCACAGTGTGCCAGCAGCTGAGAGC  644

Qy     487 TCGGCGGGTGCTCTCGCTCGTCGTCCTCGTACCTCCAAGAGTCCCAAGATCCTCAAGACCACCACCTACT  546
           |||||||||||| |||||| ||| ||||| || |||||||||| ||||||||||| ||||||||||||||
Db     645 TGGGCGGGTGCTCTCGCTCGTCGTACCTCGTACCTCCAAGAGTCCGTGTACCTCGTGTACCTGCTTTCATCA  704

Qy     547 CCAGCGGTTCGAGCTCGGCTCGGCTCCGGTAGCCGACGGGGCCCCGGGTTGCAGGCGACCTGCGCA  606
           ||||| || ||| |  |||  |||  | ||| ||||| | |||||||||||  || ||||  ||||||
Db     705 CCAGCATGCTGAAGT---CTGCAGGCACCAATGGACCCCGGGTTCAGGGAGAGCTGAGAA  761

Qy     607 ATCTGTGCCGCCTGGCCGGGGCGGAGGGCCCTCG---GACGGCAGGAGGGGCCAGTTGCCGAGA  663
           ||| ||||||||||  |||||| ||||| ||||||   ||||| ||||  ||| ||||||| |||
Db     762 CGCTGTGTCGGCCTGGCCAGGAGGAGGGCATCCTGGATCATCATAAAGCACAATTCTCAGAAG  821

Qy     664 AGGAACTGGAGCAACTGGAGCTTCGTGGCTCAAGTGCAGACGGCTGTTCTCAGCAAAA  723
           ||||| ||||||  |||||| ||| ||||||  |||| ||| ||| |||||||||||||
Db     822 AGGACCTGGAGAAATTGAAGCTTCGGGGTTCCCCAAGTTCAGACAATATTTCCAACAGA  881

Qy     724 AGGAGCTGCCGGGCGTGCTGGAGGACAGAGGTCATCGACCAGAGCTTCC  783
           |||||  ||| ||| | ||| |  | || ||||||| ||||||||||| |
Db     892 AAGAGATACCAGGAGTGCTAAAAACTGAGGTCACCTACCAGTCATTGACCAGAGCTTCC  941
```

RESULT 2
AI344276

LOCUS   AI344276   691 bp  mRNA  linear  EST 08-APR-1999
DEFINITION tc03a07.x1 NCI_CGAP_Co16 Homo sapiens cDNA clone IMAGE:2062740 3'
      similar to TR:Q63035 Q63035 VASOPRESSIN RECEPTOR. ;, mRNA sequence.
ACCESSION  AI344276
VERSION   AI344276.1  GI:4081482
KEYWORDS  EST.
SOURCE   Homo sapiens (human)
 ORGANISM Homo sapiens
      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
      Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1  (bases 1 to 691)
 AUTHORS  NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
 TITLE   National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
      Tumor Gene Index
 JOURNAL  Unpublished
COMMENT   Contact: Robert Strausberg, Ph.D.
      Email: cgapbs-r@mail.nih.gov
      Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.
      , Ph.D.
      cDNA Library Preparation: M. Bento Soares, Ph.D.
      cDNA Library Arrayed by: Greg Lennon, Ph.D.
      DNA Sequencing by: Washington University Genome Sequencing Center
      Clone distribution: NCI-CGAP clone distribution information can be
      found through the I.M.A.G.E. Consortium/LLNL at:
      www-bio.llnl.gov/bbrp/image/image.html
      Insert Length: 1242  Std Error: 0.00
      Seq primer: -40bp from Gibco
      High quality sequence stop: 459.
FEATURES     Location/Qualifiers
 source    1..691
      /organism="Homo sapiens"
      /mol_type="mRNA"
      /db_xref="taxon:9606"
      /clone="IMAGE:2062740"
      /tissue_type="colon tumor, RER+"
      /lab_host="DH10B"
      /clone_lib="NCI_CGAP_Co16"
      /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a
      modified polylinker; Site 1: Not I; Site 2: Eco RI;
      Plasmid DNA from the normalized library NCI_CGAP_Co10 was
      prepared, and ss circles were made in vitro. Following HAP
      purification, this DNA was used as tracer in a subtractive
      hybridization reaction. The driver was PCR-amplified cDNAs
      from a pool of 5,000 clones made from the same library
      (cloneIDs 1057416-1061255, and 1144584-1145351).
      Subtraction by Bento Soares and M. Fatima Bonaldo.  "

[Sequence alignment data — Query (Qy) and Database (Db) nucleotide sequences with match bars, not transcribed in full]

BASE COUNT    100 a    232 c    245 g    113 t    1 others
ORIGIN

Query Match         20.5%;    Score 550.8;    DB 9;    Length 691;
Best Local Similarity   98.6%;    Pred. No. 6.5e-43;
Matches   555;    Conservative   0;    Mismatches   8;    Indels   0;    Gaps   0;

```
Qy      7  CGTGGGCGCAGCCTGGCTGACCTGATCCTGACCAGTGCCCGGCGGCGGCCGG  66
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129  CGGGCACGCGAGCCTGGCTGCACCTGATCCTGACCAGTGCCCGGCGGCGGCCGG  188

Qy     67  TGCCGCAGATGCTGCCCCAGCCGGCTGCTTCATCTGGCGACGGCGGCGACGAGC  126
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    189  TGCCGCAGATGCTGCCCCAGCCGGCTGCTTCACCTGGCGACGGCGGCGACGAGC  245

Qy    127  TGCCGGCGGTGGGGGGCCCCGAGGCGGGCCCTCGACAGACCCTCGAGGCGGCAGCG  186
           |||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db    249  TGCCGGCGGTGGGGGGCCCCAGGCGGGCCCTGCACAGACCCTCGAGGCGGCAGCG  308

Qy    187  GCGGCGGGTGCTAGGCGGCTGCTGAGTAAGGCGCCTGCGCCCACGGCCCTCCTGTGG  246
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db    309  GCGGCGGGTGCTAGGCGGCTGCTGAGTAAGGCGCCTGCGCCCACGGCCCTCCTGTGG  368

Qy    247  TGACCACCGGCGCGCGCCGCCCCGGGAGGCTGCAGGCGGCCGCCGCTGTGTTCCCCGCAGTGCG  306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    369  TGACCACCGGCGCGCGCCGCCCCGGGAGGC GCAGGCGGCCGCCGCTGTGTTCCCCGCAGTGCG  428

Qy    307  CCGAGGTGCGGCGGCTTCTCGCACAGGACAAGAGAAGTATTTCTACAAGTTCT CCGG  366
           ||| |||||||||||| ||| |||||||||||||||||| |||||||||||| ||||
Db    429  CCGAGGTGCGGCGGCT CTCCACAGGACAAGAGAAGTATTTCTACAAGTTCTCCGG  488

Qy    367  ATGAGAGGAGGGGGCGAGGGCGCTACCCGTTCGTGAGGAGAACGAGAGCGCTGTTCGGCGC  426
           |||||||||||||||||||||||| ||| |||||||||||||||||||||||||||||||||
Db    489  ATGAGAGGAGGGGGCGAGGGCGCCTACCCGTTCGTGAGGAGAACGAGAGCGCTGTTCGGCGC  548

Qy    427  TGTGCTTCGCTGCCCTTCGTCGCTGGATCGTGTGCCACCGTGGTGCTGCGCCAGCAGCTGGAGC  486
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    549  TGTGCTTCGCTGCCCTTCGTCGCTGGATCGTGTGCCACCGTGCGCAGCAGCTGGAGC  608

Qy    487  TCGGTCGGGACCTGTCGCGCGACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCATCA  546
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    609  TCGGTCGGGACCTGTCGCGCGACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCATCA  668

Qy    547  CCAGCGGTTCTGAGCTCGGCTCCG  569
           ||||||||||||||||||||||||
Db    669  CCAGCGGTGCTGAGCTCGGCGCG  691
```

RESULT 3
AI762985/c
LOCUS       AI762985        427 bp    mRNA    linear    EST 24-JUN-1999
DEFINITION  wi05e11.x1 NCI_CGAP_CLL1 Homo sapiens cDNA clone IMAGE:2389388 3',
            mRNA sequence.
ACCESSION   AI762985
VERSION     AI762985.1    GI:5178652
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 427)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Asn Alizadeh, John Byrd, M.D., Mike Grever,
            M.D., Louis M. Staudt, M.D., Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be

found through the I.M.A.G.E. Consortium/LLNL at:
www-bio.llnl.gov/bbrp/image/image.html
seq primer: -40UP from Gibco
High quality sequence stop: 422.

FEATURES             Location/Qualifiers
     source          1..427
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2389388"
                     /tissue_type="B-cell, chronic lymphotic leukemia"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_CLL1"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site_1: Not I; Site_2: Eco RI; 1st strand cDNA
                     was primed with a Not I - oligo(dT) primer 5'
                     TGTACCAATCTGAAGTGAGACGGCCGCATTGCTTTTTTTTTTTTTTTTTTTT
                     T 3'; double-stranded cDNA was ligated to Eco RI
                     adaptors (Pharmacia), digested with Not I and cloned into
                     the Not I and Eco RI sites of the modified pT7T3 vector.
                     Library is normalized, and was constructed by Bento
                     Soares and M.Fatima Bonaldo."

BASE COUNT     83 a    108 c    130 g    106 t
ORIGIN

Query Match         15.3%;    Score 411;    DB 9;    Length 427;
Best Local Similarity   97.7%;    Pred. No. 9.4e-30;
Matches   417;    Conservative   0;    Mismatches   10;    Indels   0;    Gaps   0;

```
Qy   2210  AGACCCTCCAGCCATTCCCCTGAAAAACACCCCCAAGCCACCAAGCCACAATAATGACAGC  2269
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    427  AGACCCTCCAGCCATTCCCCTGAGAACACCACCACCAGCCACGACCCAAGCCTCAATAATGACAGC  368

Qy   2270  GAGAGCTTCAATTAACTAAGCACCTACCTGGGCGCAGAATAACCCTTCACTGCCTGATCC  2329
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    367  GAGAGCTTCAATTAACTAAGCACCTACCTGGGCGCAGAATAACCCTTCACTGCCTGATCC  308

Qy   2330  CCATCTGCAGTGTGGCCCAACAGCCCCCAGACTATGCCCATAGACTGGAGGTAGGCA  2389
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    307  CCATCTGCAGTGTGGCCCACCAGCCCCCAAAGAATTATGCCCACATAGACTGGAGGTAGGCA  248

Qy   2390  GTTCACCGTCCCCTGTTAGGAATGAGACCATCCTGAGGCTATGGCCCAGGCCCACA  2449
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    247  GTTCACCGTCCCCTGTTAGGAATGAGACCATCCTGAGGCTATGGCCCAGGCCCACA  188

Qy   2450  GGGGTCCAGTGTCTGAGATCTTTGGGAAGGGAGACTAGGGCAGGTCAGGTGGAGACAGCGCAGGA  2509
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187  GGGGTCCAGTGTCTGAGATCTTTGGGAAGGGAGACTAGGGCAGGTCAGGTGGAGACAGCGCAGAA  128

Qy   2510  CCCCCGTGCTGGGTGGGAAGCATGACCACATGGTGGGTGAGCAGCCCCCATGCCACTGACG  2569
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    127  CCCCCGTGCTGGGTGGGAAGCATGACCACATGG?GGGTGAGCAGCCCCCATGCCACTGACG  68

Qy   2570  GTAAATTCCCCTGTGGACTCATTTCTGTGGTTTCTATTACACCTGCGGTGTGTAC  2629
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     67  GTAAATTCCCCTGTGGACTCATTTCTGTGGTTTCTATTACACCTGCCAGAAGTAGTAC  8

Qy   2630  AATACAG  2636
           |||||||
Db      7  AATACAG  1
```

RESULT 4
AW107005
LOCUS       AW107005        798 bp    mRNA    linear    EST 20-OCT-1999
DEFINITION  um18d07.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2192653 5' similar to TR:Q63035 Q63035 VASOPRESSIN RECEPTOR.
            ;, mRNA sequence.
ACCESSION   AW107005
VERSION     AW107005.1    GI:6077873
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus

```
Qy   467 GCTGGCGCCAGCAGCTGGAGCTCGGTCGGAGGACCTGTCGCGCACGTCCAAGACCACCAGTC 526
             ||||| ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   300 GCTGCAGCAGCAGCTAGAGCTGGGCCGGGATCTCTCTCGTACCTNCAAGACCACTACTTC 359

Qy   527 AGTGTACCTGCTTTCATCACCAGCGTTCTGAGCTCGGTCTCCGGTAGCGACGGGCCCCG 586
             ||||||||||||||||||||||||||||||||
Db   360 TGTGTACCTGCTCTTCATCACCAGCATGCTGA---AGTCTGCAGGCACCAATGGACCCCG 416

Qy   587 G:TGCAGGGCGACCTGGCCAATCTGTGCGCCTGGCCCGCGGAGGGGGCTCCTCGGACGC-- 644
             ||| |||||||||||||||||||||||||||||||||| ||||||| ||||||||||||
Db   417 GGTTCAGGGAGAGCTGAGAACGCTGTGT:CGCCTGGCCCGCCAGGGAGGGGCATCCTGGATCATCA 476

Qy   645 -AGGGGCGAGTTGCCCGAAGAAGAAGCTGGAGCAACTGGAGCTTCGTGGCTCCAAGTGCA 703
             | ||||||||||||||||||||||||||||||||||
Db   477 TAAAGCACAATTCTCAGAAGAGACCTGGAGAAATTGAAGCTTCGGTGTTCCCAAGTTCA 536

Qy   704 GACGCGTGTTTCAGCAAAAAGGAGCTGCCGGCGTCTGGAGACAGAGGTCACCACCA 763
             ||| ||||||||||||||||||||||||||||||||
Db   537 GACAATATTTCTCAACAAGAAAGATACCAGGAGTGCTAAAAAC:GGAGTCACCTACCA 596

Qy   764 GTTCATCGACCAGAGCTTCCAGGAGTTCCTCCGGGACACTGTCCTCACCTGCTGGAGGACGG 823
             |||||||||||||||||||||||||||||||||||
Db   597 GTTCATTGACCAGAGCTTCCAGGAGTTCTTGGCTGCATTGTCATTGTCATACCTAC:AGAGAGCTGA 656

Qy   824 CGGGGTGCCCAGGACGCGGGCGTGCGGCGGCGTTGGGACACTCCTGCGTGGGGACGCCCAGCC 883
             ||| |||||||||||||||||||||||||||||||
Db   657 GCGAACTCCGGGGACGCCGTCGTCAGCGGTGTGCAGAAGCTCCTGAACTACTACGACGCGAGC: 716

Qy   884 GCACAGCCACTTGGTGTCTCACCACGCGCTTCCTCTT 919
             |||||||||||||||||||||||||||||||
Db   717 ACG:GGTCATCTCTGCACTGACCACCCGATTTCTCTT 752
```

```
RESU:1.5
BG971110
LOCUS       BG971110        719 bp    mRNA      linear     EST 12-JUN-20C1
DEFINITION  602837817F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4971940
            5', mRNA sequence.
ACCESSION   BG971110
VERSION     BG971110.1  GI:14358747
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 719)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAMI0958  row: b  column: 05
            High quality sequence start: 6
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4971940"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Kid14"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
```

```
REFERENCE   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 798)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished
COMMENT     Other ESTs: um18d07.xi
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1105105
            Seq primer: custom primer used
            High quality sequence stop: 503.
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2192653"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse kidney mxia"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraII
                     (CACTGTGTG); Site_2: DraII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     (ATGTGGCCTTTTTTTTTTTTTT); double-stranded cDNA was
                     ligated to a DraII adapter (TGTTGGCCTACTGG), digested
                     and cloned into distinct DraII sites of the pME18S-FL3
                     vector (5' site CACTGTG, 3' site CACCATGTG). XhoI should
                     be used to isolate the cDNA insert.  Size selection was
                     performed to exclude fragments <1.5kb. Library
                     constructed by Dr. Sumio Sugano (University of Tokyo
                     Institute of Medical Science).  Custom primers for
                     sequencing: 5' end primer CTTCTGCCTAAAAGCTGCG and 3' end
                     primer CGACCTGCAGCTCGACACA."
BASE COUNT      190 a     203 c     222 g     182 t     1 others
ORIGIN
```

```
Query Match.      15.0%;    Score 403.8;  DB 9;  Length 798;
Best Local Similarity.  73.5%;   Pred. No. 2.9e-29;
Matches 556;  Conservative  0;  Mismatches 193;  Indels 7;  Gaps  3;

Qy   167 CCCCTCGAGGCGCGAGCGCGCGGGCTCGTAGGCGGGCTGCTGAGTAAGGCGCTGCT 226
             ||| ||  |||||||||||||||||||||||||||
Db     1 CCCTTGAGGCCACGAGTGGCTTGAGAGTGCTGAGCGGCTGCTGAGCGCCAGGAGCTGTT 60

Qy   227 GCCCACGGCGCCTCCTGCTGGTGACCACGCGGCGCGCCCCGGAGGCGCTGCAGGGGCCG 286
             ||||||||||||||||||||||||||||
Db    61 ACTGTA-GCTCGCTTGCTAGTGACTACACGCCAGGCCGTTACACGGGAGGGGGCTGCAGGG 119

Qy   287 CCTGTGTTCCCCCAGTGCGCCGAGGTGCGCCGGCTTCCCGACAAGGACAAGAAGAAGTA 346
             |||||||||||||||||||||||||||||||
Db   120 ATTGTGCCTCGCCACAGTGCGCGAAATACGCGGCTTCTCAGACAACAAGACAAAAAAAGTA 179

Qy   347 TTTCTACAAGTTCTTCCGGGATGAGAGGAGGGGCCGAGCGCGCCTTCGTGAAGGA 406
             ||||| |||||||||||||||||||||||||||
Db   180 TTTCTTCAAGTTCTTCCGGGACGAGTGGCTTGAGAGTGCTGAGCGCGCGTTACCCGCTTCCGTTCGTGAAAGA 239

Qy   407 GAACGAGAGCGCTGTTCGCGCCTGTGTCTTCGTGCCTTCGTGCTGATCCGCGTGCACCGT 466
             ||||||||||||||||||||||||||||||||
Db   240 GAACGAGAGACCGCTGTTCACTCGTGCTTTGCACTCGCGTTTGTGCCCTTCGTGTGCTGGATCGTGTGCACAG: 299
```

COMMENT

Contact: Robert Strausberg, Ph.D.
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10959  row: h  column: 22
High quality sequence stop: 601.

FEATURES             Location/Qualifiers
     source          1..603
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4972485"
                     /clone_lib="NCI_CGAP_Kid14"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT
ORIGIN

BASE COUNT      :76 a      189 c      205 g      149 t

Technologies. Note: this is a NCI_CGAP Library. |"

Query Match          14.4%;   Score 386;  DB 12;  Length 719;
Best Local Similarity  75.9%;  Pred. No. 1.4e-27;
Matches 530;  Conservative  0;  Mismatches 160;  Indels  8;  Gaps  4;

[sequence alignment data]

RESULT 6
BG971489
LOCUS       BG971489               603 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602838184F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4972485
            5', mRNA sequence.
ACCESSION   BG971489
VERSION     BG971489.1  GI:14359125
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 603)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished

```
Db   599 TACCA 603

RESULT 7
BF783992
BF783992
LOCUS       BF783992          938 bp    mRNA    linear    EST 12-JAN-2001
DEFINITION  6021078335F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236272
            5', mRNA sequence.
ACCESSION   BF783992
VERSION     BF783992.1  GI:12089028
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 938)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM9844  row: e  column: 09
            High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source          1..938
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236272"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Kid14"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT    262 a    232 c    2?4 g    170 t
ORIGIN

Query Match       12.5%;  Score 337.4;  DB 10;  Length 938;
Best Local Similarity   76.0%;  Pred. No. 3.9e-23;
Matches  507;  Conservative  0;  Mismatches  151;  Indels  9;  Gaps  7;

Qy   221 GCTGCTGCCCACGGCCCCTCCTGCTGGTGACCACGCGCGCCCCGCGCCCCGGGAGGCTGCA 280
             ||||   || ||||||||| |||| | ||| |   ||| |   ||  ||||||
Db     2 GCTGT-ACCAGGAGCTCGCTTCCTCCTAGTGACTACACGCCATGCGCTACACGGGAGGCTGCA  61

Qy   281 GGGCCGCCTGTGTTCCCGCAGTGCGCCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAA 340
             |||||||||||| ||||||||||  |||||||||| ||||||||||||||||||||||
Db    62 GGGCAGA-TGTGCTCGCCACAGTGCGCAGAAATACGCGGCTTCTCAGACACAAAGACAAAAA 120

Qy   341 GAAGTAGTTCTACAAGTTCTTCCGGGATGAGAGGAGGAGGCCGAGCGGGCCTACCGCTTCG? 400
             || ||||||  ||||||||||||  |||||||||||||||  |||||||  ||||||
Db   121 AAAGTATTTCT?CAAGTTCTTCCGGGACGAGGAGGAGGCGGCGCT?ACCGCTTCGT 180

Qy   401 GAAGGAGAACGAGACGCTGTTCGCCGCTGTGCTTCGTGCGCTGGATCGTGTG 460
             || ||||||||||||||||||||  ||||||| |||||||||||||||||||
Db   181 GAAAGAGAACGAGACGCGCTGTT-GCACTGTGCTT-GTGCCCTTCGTGCGCTGGATCGTGTG 238

Qy   461 CACCGTGCTGCGCCACAGCTGAGCTCGT?CGGGACTCGTCGCCACGTCCAAGACCAC 520
             ||||||||||||| ||||||||||||||||  ||||||||||||||||||||||||||
Db   239 CACAGTGCTGCAGCAGCAGCTA?AGCTGGCCCGGATCTCTCGTACCTCCAAGACCAC 298

Qy   521 CACGTCAGTGTACCTGCTTTCATCACCAGCCTCTGAGCTCGGCTTCGGTAGCCGACGG 580
             ||||||||||||  ||||||||||| |||||||||||||||  ||||||||||
Db   299 TACTTCTGTGTACCTGCTCTTCATCACCAGCATGCTGAAG-CTTCTGCAGCCACCAATGG 357
```

```
Qy   581 GCCCGGTTGCAGGGGCGACCTGCGCAATCTGTGCCGCCTGGCCGCGCAGGGGGTCCTCG- 639
             ||||| |||||||||||||||||  ||||| ||||| |||||||| |||||||||||||
Db   358 ACCCGGGTTCAGGGAGCTGAGAGCTGAGAACGCTGTGTCGTCGTCGGCCCTGGCGCAGGGGCATCCTGGA 417

Qy   640 --GACGCAGGGGCAGTTGCCGAGGAGAACTGGAGCAACTGGAGCTTCGTGGCTCCAA 697
               |||||||||||||  |||||| |||| ||||||  ||| ||||| |||||||||
Db   418 TCATCATAAAGCACAATTCTCAGAAGAGGACCTGGAGAAATTGAAGCTTCGGGGTTCCA 477

Qy   698 AGTGCAGACGCCGTCTTCTCAGCAAAAAGGAGGTGCCCGGCCGTGCT-GGAGACAGAGGTCA 756
             ||| |||| ||||||||||||| |||||||||||||||| | |||||| ||||||||||||
Db   478 AGTTCAGACAACAATATTTCTCAACAAGAAGAGATACCAGGAGTGCTAAAAAAACTGAGGTCA 537

Qy   757 CCTACCAGTTCATCGACCAGAGCTTCCAGGAGTTCCTCGCGGCACTCGTCCTACCTCTGG 816
             ||||||||||||||| ||||||||||||||||||| ||||||||||| ||||| ||||
Db   538 CCTACCAGTTCATTGACCAGAGCTTCCAGGAGTTCTTGGCTGCATTGTCATACCTACTAG 597

Qy   817 AGGACG-GCGGGGGTGCCCAGGACCGCGCTGGCGGCGCTGGGACACTCCTGCGTGGGGAC 875
             ||||||  |||||||||||||||||||| |||| ||||||| |||||||||||||||||
Db   598 AAGCTGAGCGAACTCCCGGGACACGCGCCCCAGGCGCGTGTGCAGAAGCTCCTGAACTCTGAC 657

Qy   876 GCCCAGC 882
             ||   ||
Db   658 GCCGAAC 664
```

```
RESULT 8
BB633515
BB633515
LOCUS       BB633515  RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330055K17 5', mRNA sequence.          670 bp    mRNA    linear    EST 26-OCT-2001
DEFINITION  BB633515 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330055K17 5', mRNA sequence.
ACCESSION   BB633515
VERSION     BB633515.1  GI:16469948
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
AUTHORS     Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,A., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,T., Sogabe,Y., Suzuki,H.,
            Tagami,M., Tagawa,A., Takahashi,F., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       RIKEN Mouse ESTs (Arakawa,T., et al. 2001)
JOURNAL     Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center (GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/,
            Carninci,P. Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res.  10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T. Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
```

nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and Hayashizaki,Y.
Computational Analysis of Full-Length Mouse cDNAs Compared with Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
Please visit our web site (http://genome.gsc.riken.go.jp) for further details.

```
FEATURES             Location/Qualifiers
     source          1..670
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /db_xref="taxon:10090"
                     /clone="A330055K17"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="RIKEN full-length enriched, adult male spinal
                     cord"
                     /note="Site 1: SalI; Site 2: BamHI; cDNA library was
                     prepared and sequenced in Mouse Genome Encyclopedia
                     Project of Genome Exploration Research Group in Riken
                     Genomic Sciences Center and Genome Science Laboratory in
                     RIKEN. Division of Experimental Animal Research in Riken
                     contributed to prepare mouse tissues. 1st strand cDNA was
                     primed with a primer [5'
                     GAGAGAGAAGGATCCAAGAGCTC^TTTTTTTTTTTTTTVN 3']. cDNA was
                     prepared by using trehalose thermo-activated reverse
                     transcriptase and subsequently enriched for full-length by
                     cap-trapper. cDNA went through one round of normalization
                     to Rot = 10.0 and subtraction to Rot = 459.0. Second
                     strand cDNA was prepared with the primer adapter of
                     sequence [5' GAGAGAGATTCGAGTTAATTAATTAATCCCCCCCCCCCCCCC
                     3']. cDNA was cleaved with XhoI and BamHI. Vector: a
                     modified pBluescript KS(+) after bulk excision from Lambda
                     FLC I."
BASE COUNT     143 a      189 c      210 g      128 t
ORIGIN

Query Match        11.8%;  Score 316.4;  DB 10;  Length 670;
Best Local Similarity   77.9%;  Pred. No. 4.4e-21;
Matches   394;  Conservative    0;  Mismatches  111;  Indels    1;  Gaps    1;

Qy     7 CGTGGGGCGGCCAGCCTGCGTGACCTGATCCTGGACCAGTGCCCGACCCGGCGGCGGCCGG     66
         ||||    ||||||| |   ||||||| ||||||||||||||||||  |||| | || ||||
Db   165 CGGGAAAGGCCAGCTTGGGGGACCTGGTCTGGACCAGTGCCCGGGACCGCGCTTGGCCAG    224

Qy    67 TGCCGCAGATGCTGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGGACGAGC    126
         ||  || |||| | || | | ||| |||| || |||||||| | |||||| | ||||||
Db   225 TGAAGCGGATACTGCGCAACCCAATCGCCTACTGTTCATCCTCGGACGGTGCAGACGAGC    284

Qy   127 TGCCGGCGGCTGGGGGGCCCCGAGGCGCGGGCCCTGCCACAGACCCCTTCGAGGCGGCGAGCG    186
         ||||  | ||| |||    ||||||||  | ||||| |||| |||||| | |||| || || |
Db   285 TGCCTACTTTACCATCCTCAGAGAGGCCACTCCCTGCAAAGACCCTTTGGAGGCTACAAGTG    344

Qy   187 GGGCGGCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGGCCCTGCGCCCAGGGCCCTCCTGCTGG    246
         | ||| ||| |||||||| |||||||||| ||||| || |||  ||||||  ||| || ||||
Db   345 GCTTGAGAGTGCTGAGCGGCGGCTGCTGAGCCAGGAGCGTGTTACCAGGAGCTCGGTTGCTAG    404

Qy   247 TGACCACGGCGCGCGCCCCGGGAGGCTGCAGGGCGCCGCCTGGGTTCCCGCAGTGCG    306
         |||| || |||  |||| | |||||  ||||| || || || |||||||| ||||||
Db   405 TGACTACACCGCCATGCCGCTCACAGGGAGGAGGCTGCAGAGTTGTGCTCGCCACAGTGCG    464

Qy   307 CCGAGGTGCGCGCTTCCGCACAGGACAGAGAGAAGTATTCTCACAAGTTCTTCCGGG    366
         | ||| || || ||| | ||||| || || | | ||||||| | |||||||||| ||
Db   465 CAGAAAACCGCGGCTTCTCAGACACAGACAAAAAAAGTATTCTTCTTCAAGTTCTTCCCGGG    524

Qy   367 ATGAGGAGGAGGGCGCGAGGGGCCGAGGGCGCGCCTACCGCTTCGTGAAGGAGGAAGAACGCGTGTTCGCGC    426
         | |||||||| | |  |||| || || | |||| ||| | |||||||||  ||  | ||||| ||||| | |
Db   525 ACGAGGAGGAAGGCAGAGGCGCAGCGCGCCCTACCCCACCGCGTTCGTGAAAGA-AACGAGACGCGTGTTGCAC    583

Qy   427 TGTGCTTCGTGCCCTTCGTGTGCTGATCGTGCCGTGCGGCCAGCAGCTCGGAGC    486
         |||||||  ||||||||| || |||||||||| | |||| ||| ||||| ||||
Db   584 TGTGCTTTGTGCCCTTCGGGTTGCTGGATCGTGTGCGACAGTGCTGCAGCAGCTAGAGC    643

Qy   487 TCGGTGGGACCTGTCGCGGCACGTCC    512
         | ||| ||| ||  |||   ||| ||
Db   644 TGGGGCCGGGGTCTCTTCGTACATCC    669
```

RESULT 9
BG965959

```
LOCUS       BG965959                 832 bp    mRNA    linear   EST 12-JUN-2001
DEFINITION  602829751F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4984429 5',
            mRNA sequence.
ACCESSION   BG965959
VERSION     BG965959.1  GI:14353596
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 832)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10990 row: j  column: 14
            High quality sequence stop: 676.
FEATURES             Location/Qualifiers
     source          1..832
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4984429"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Co24"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size .6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP library."
BASE COUNT     225 a      211 c      243 g      153 t
ORIGIN

Query Match        11.4%;  Score 307.6;  DB 12;  Length 832;
Best Local Similarity   69.3%;  Pred. No. 2.5e-20;
Matches   538;  Conservative    0;  Mismatches  219;  Indels   19;  Gaps    8;

Qy   510 TCCAAGACCACCACCAGTCAGTCAGTGCTACCTGCTTTCATCACCAGCGTTCTGAGCTGCGGCCCCG    569
         |||||||||| | | | ||| | ||||| ||||| |||||||||  ||| |||||  || |||
Db     1 TCCAAGACCACTACTTCTGTGTACCTGCTCTTCATCACAAGCATGCTGAAGTCTGCA----    57

Qy   570 GTAGCCACGGCGCCCCCGGTTGCAGGGCGACCTGGCCAATCTGTGCCGCCCGCCCCGGAG    629
         | | |||   | ||||||| | |||  | || | |  | | | |||| | |||| |||||
Db    58 GGCACCAATGGACCCCCGGGTTCAGGAAGAGCTGAGAACGGTCGTGTCGTCGCCTGGCCAGGGAG    117

Qy   630 GGGCGTCCTCG---GACGCAGGCGCAGTTGCCGAGGAAGGAACTGGAGCAACTGGAGCTT    686
         || | ||| ||   || || || | ||  | | ||||   | |||||| | |||||||||
Db   118 GGCATCCTGGATCATCATCAAAGACCACAATTCTCAGCAGGACCTTGAGAGAATTGAAGCTT    177

Qy   687 CGTGGCCCAAAGTGCAGACGACCCGGCTGTTTCTGCAGACAGTGCTGGAG    746
         || || ||||| | ||| |||  | || ||||||| | |||  | |||||
Db   178 CGGGGTTCCCAAGTTCAGACAATATTTCTCAACAAGAAAGATACCAGGAGTGCTAAAA    237

Qy   747 ACGAGGTCACCTACCACCAGTTCATCGACCAGAGCTTCCAGGGGTTCCTCGGCGCACTGTCC    806
```

```
Db  238 ACTGAGGTCACCTACCAGTTCATTGACCAGAGCTTCCAGGAGT-CTTGGCTGCATGTCA  297
          || |  |||||||||  |||||||||||  |||||||||||||| | | ||||||

Qy  807 TACCTGCTGGAGGACGGCGGGGTGCCCAGGACCGGCCGTGGGAGTCCTGGGACTCCTG  866
          ||||||||||||  ||||| |||| |||

Db  238 TACCTACTAGAAGTGAGCGAACTCCGGGACGCCCAGGCGGTGTGCAGAAGCTCCTG  357
          ||||||||||||| |||  |||

Qy  867 CGTGGGGACGCCCAGCCGACCCACAGCCCACTTGGTGCTCACCACGCGC-TCCTCTCGGACTG  926
          ||||| ||||  ||||||||||| |||  |  ||||  ||||||

Db  358 AACTCTGACGCGGAGCTACGTGGTCATCTTGCACTGACCACCCGATTCCTC-TTGGCCTG  417
          ||||| |||||||

Qy  927 CTGAGCCGCGGACCGGATGCGGCGGACATCGAGCGCCACTTCGGCTGCTCCATGGTTTCAGAGCGT  986
          ||||||||||||||||||||| |||| ||||||||||||||||||||||||| |||

Db  418 CTAAATACGAGAGGGGCTTCGTGTGACATTGGAAACATTTTGCTGTTGGTGCCAGATCAT  477
          ||||||||||| ||| ||||

Qy  987 GTGAAGC-AGGAGGGCCTGCGGTGGGTGCAGGAGGACAGGGAGGACAGGGCTGCCCCGGAGTGGC 1045
          |||||||||||||| ||||| ||||||||

Db  478 GTGAAACAAGGACACCCTGCGTGGGT------ACAAGGACAGAGCCACCCAAGGGGCC  531
          ||||||||||||||

Qy 1046 ACCAGAGGTGACCGAGGGGCCAAAGGGCTCGAGGACACCCAAGAGCCAGAGGAGGAGGA 1105
          || ||||||||||| ||||||||| |||| ||||| |||||

Db  532 ACCAGTAGGGGGCAAAAAAGACTGCTGAGCTGGAGGCATCCGAGGACGCAGGAGGAGGA  591
          ||||||||| |||||

Qy 1106 GGAGGGAGAGAG---CCCAATACCCACTGGAGTTGCTGTACTGCCTGTACGAGACCGCA 1162
          |||||||||||||||||||| |||| |||||

Db  592 GGAGGAAGAGGAGGACCTCAACTCCGGACTGGAGCTGCTGTACTGCCTGTATGAGACACA  651
          |||||||| ||

Qy 1163 GGAGGACGCGTTTGTGCGCCAAGCCTCCGTGTCCGGGTTCCCGGACGCTGGCGCTGCAGCGAGT 1222
          ||||||| |||||||| |||||||

Db  652 GGAGGACGGATCAGT-CGACAGCGGGATCGTGTTCCAGCAGCCTTCCAGAGATAGTACTGGGAGCCGAGT  720
          |||||||||||||||| |||

Qy 1223 GC-GCTTCTGCCGCATGG-ACGTGGCTGTTCTGAGCTACTGCGTGAGG-GCTGCCC 1275
          | ||||||||| ||||  ||||||

Db  711 CCAGGTTGAGCCGCCGCCATGAACCTTAGAACGTCTGAATTACGGCATGCAGCGGATGCCC  766
          |||||
```

---

```
RESULT 10
BY262618
LOCUS       BY262618                    416 bp    mRNA        linear        EST  10-DEC-2002
DEFINITION  BY262618 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K330i4H14 5', mRNA sequence.
ACCESSION   BY262618
VERSION     BY262618.1   GI:26444130
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 416)
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,R., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            ,Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
            ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
```

---

```
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
MEDLINE     22354683
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000)
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (11), 1757-1771 (2000)
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
            cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Tissues were provided by Michela Fagiolini and Takao K. Hensch;
            Laboratory for Neuronal Circuit Development Brain Science Institute
            RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
            assistance we gratefully acknowledge. Please visit our web site
            (http://genome.gsc.riken.go.jp) for further details.
FEATURES             Location/Qualifiers
     source          1..416
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K330i4H14"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT      94 a    109 c    126 g     87 t
ORIGIN
Query Match     10.0%;  Score 268.8;  DB 13;  Length 416;
Best Local Similarity   79.6%;  Pred. No. 1.7e-16;
Matches  331;  Conservative   0;  Mismatches  82;  Indels   3;  Gaps  1;
```

```
Qy  191 GCGGTGCTGGGCGGGGCTGCTGCTGCTAGGTAAGGCGCTGCCCACGGCCCCTCCTGTGTGAC  250
          ||||||||||||||||||||| |||||||||||||||| |||||| ||||||||||||||

Db    2 GAGAGTGCTGAGCGGGGCTGCTGAGCCAGGAGAGCTGTTACCAGGAGCTCGCTGCTAGTGAC  61
          |||||||||||||||||| |||||| ||| ||||| ||| |||||||||||||| ||||||

Qy  251 CACGGGCGGCCCGCCCCGGAGGGTGCAGGGCTGCAGGGCAGCAGATTGTGCTCGGCAGTGCGCCGA  310
          |||||||||||||||||||| |||||||||||||||||||||||||||| |||||||||

Db   62 TACACGCCATGCGGTCACAGGGCAGGGCAGGCTGCAGGGCAGATTGTGCTCGGCCACAGTGGCCAGA  121
          ||||||||||||| ||| |||||||||||||| |||||||||||

Qy  311 GGTGCGGCGGCCTTCTCCGACAAGGACAAGCAAGACAAGCAGAGAAGAGTATTTCTACAAGTTCTTCCGGGATGA  370
          |||||||||||||||||||||||| |||||||||||| |||||||||||||||||||||||

Db  122 AATACGCGGCTTCTCAGACAACAAAGACAAAAAAGTATTTCTTCAAGTTCTTCCGGGAACGA  181
          |||||||||||||||||||||||| |||||||||||||||||||||

Qy  371 GAGGAGGCGGGAGCCGCCGCTACCGCTTCGTGAAGAAGGCAACAACAACGCGTGTTCCGCGCTGTG  430
          ||||||||||||||||||||||||||||| ||||
```

```
Db   182  GAGGAAGGCAGAGCGGCGCTTACCGTTCGTGAAAGAGAAGAACGAGACGCGTGTTTGGACTGTG  241

Qy   431  CT-CGTGCCTTCGTGTGCTGGGATCGTGTGCACCGTGCGCCAGCAGCTGCGAGCTCGG  490
          | | |||||||| |||||||||||||||||||| ||||||| |||||||| |||||||
Db   242  CTTTGTGCCTT-CGTGTGCTGGGATCGTGTGCTGGATCGTGTGCACAGTGCTCAGCAGCTAGAGCTGGG  301

Qy   491  TCGGACCGTGTCGGCGCAGTCCAGAGCACCAGTCACCAGTGTACCTGCTTTCATCACCAG  550
          |||| || |||||||| |||||||||||||||||||||||||| |||||||||||||||
Db   302  CCGGGA-CTCTC-GTACCTCGTCAGACCAC-ACTTCTGTGTACCTGCTCT-CATCA-CCAG  361

Qy   551  CGTTCTGAGCTCGGTCCGTAGCCGACGGCCCCGGTTGCAGGGCGACCTGCGCA  636
          | |||||||||||||||||||||| |||||||||||||||||||| |||||
Db   362  CATGCTGA--AGTCTGCAGGCACCAATGACCCCGGGTTCAGGGAGCTGAGAA  414
```

RESULT 11

BF561521/c

LOCUS       BF561521            394 bp    mRNA     linear     EST 22-DEC-2000
DEFINITION  UI-R-CO-hw-e-04-0-UI.r1  UI-R-CO Rattus norvegicus cDNA clone
            UI-R-CO-hw-e-04-0-UI 5', mRNA sequence.
ACCESSION   BF561521
VERSION     BF561521.1  GI:11671255
KEYWORDS    EST.
SOURCE      Rattus norvegicus (Norway rat)
ORGANISM    Rattus norvegicus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
            Rattus.
REFERENCE   1  (bases 1 to 394)
AUTHORS     Bonaldo,M.F., Lennon,G. and Soares,M.B.
TITLE       Normalization and subtraction: two approaches to facilitate gene
            discovery
JOURNAL     Genome Res. 6 (9), 791-806 (1996)
MEDLINE     97044477
PUBMED      8889548
COMMENT     Contact: Soares, MB
            Coordinated Laboratory for Computational Genomics
            University of Iowa
            375 Newton Road, 4156 MEBRF, Iowa City, IA 52242, USA
            Tel: 319 335 8250
            Fax: 319 335 9565
            Email: bento-soares@uiowa.edu
            cDNA Library Preparation: M.B. Soares Lab Clone distribution:
            clones will be available through Research Genetics (www.resgen.com)
            This clone is available through the I.M.A.G.E. Consortium at
            LLNL (info@image.llnl.gov). IMAGE ID= 1781920
            Seq primer: M13 Forward.
FEATURES             Location/Qualifiers
     source          1..394
                     /organism="Rattus norvegicus"
                     /mol_type="mRNA"
                     /strain="Sprague-Dawley"
                     /db_xref="taxon:10116"
                     /clone="UI-R-CO-hw-e-04-0-UI"
                     /dev_stage="adult"
                     /lab_host="DH10B (Life Technologies)"
                     /clone_lib="UI-R-CO"
                     /note="Vector: pT7T3D-Pac (Pharmacia) with a modified
                     polylinker; Site 1: Not I; Site 2: Eco RI; The UI-R-CO
                     library is a subtracted library derived from the UI-R-A1
                     and UI-R-E1 libraries. The UI-R-A1 library consisted of a
                     mixture of individually tagged normalized libraries
                     constructed from rat placenta, adult lung, brain, liver,
                     kidney, heart, spleen, ovary, and muscle. The UI-R-E1
                     library consisted of a mixture of individually tagged
                     normalized libraries constructed from 8, 12 and 18-day
                     embryo. The tag is a string of 3-5 nucleotides present
                     between the Not I site and the oligo-dT track which
                     allows identification of the library of origin of a clone
                     within the mixture. The subtracted library (UI-R-CO) was
                     constructed as follows: PCR amplified cDNA inserts from a
                     pool of UI-R-A1 and UI-R-E1 clones from which 3' ESTs had
                     been derived was used as a driver in a hybridization with
```

the pooled UI-R-A1 and UI-R-E1 library in the form of
single-stranded circles. The remaining single-stranded
circles (subtracted library) was purified by
hydroxyapatite column chromatography, converted to
double-stranded circles and electroporated into DH10B
bacteria (Life Technologies) to generate the UI-R-CO
library. This procedure has been previously described
(Bonaldo, Lennon and Soares, Genome Research 6: 791-806,
1996)"

```
BASE COUNT   87 a  117 c   93 g   97 t
ORIGIN
Query Match         7.5%;    Score 201.2;   DB 10;   Length 394;
Best Local Similarity  72.6%;   Pred. No. 3.4e-10;
Matches  260;   Conservative   0;  Mismatches  98;  Indels  0;  Gaps  0;

Qy   670  TGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCGTGTTTCTCAGCAAAAAGGAGC  729
          ||||||| || ||||||| |||||| |||||||||||||||||||||||||||||||||
Db   394  TGGAGACATTGAAGCTTCAGGGTTCCCAAGTTCAGACAATGTTTCTCAGCAAGAGGAGC  335

Qy   730  TGCCGGGCGTGCTGGAGACAGAGGTCACCTACCACCAGTTCATCGACCAGAGTTTCCAGGAGT  789
          |||||||||||||||||||| ||||| |||| ||||||||||||||| |||||||||||||
Db   334  TGCCAGGAGTGCTAGAAACTGTGTCACCTACCAGTTCATTGACCAGAGCTTCCAGGAGT  275

Qy   790  TCCTCGGGGCACGTCCTACCTGCTGGAGGACGGCGGGTGCCCAGGACACCGGCGGTGGCG  849
          || || ||| ||| ||||| ||| || ||||||||||||||||| ||||||||||||||
Db   274  TCTTGGCTGCATTGTCATACCTACTAGACGCGTGAGGAGCCCAGGGAACTCCGGCAGGAA  215

Qy   850  GCGTTGGGACACTCCTGCGTGGGGACGGCCCACAGCCCACTTGGTGCTGCTCACCACGC  909
          |||||||||| ||||||| ||| |||||||||||||||||| |||||||||| ||| |
Db   214  GTGTGCAGATGCTCCTGAACTCGACCGGGCGGCTGCGTGGTCATCTGCACTCACCACTA  155

Qy   910  GCTTCCTTCCTGGACTGCTGGACGCGCTGGAGCGGATGGCGGACATCGAGCGCACTTCGGCT  969
          |||||||||| |||||| || |||||| ||||| |||||| |||||||||| ||||||
Db   154  GATTCCTTCTTGGACTGCTAAGTACAGAGAGAGGATTCGTGACATTGGAAACCATTTTGGC  95

Qy   970  GCATGGTTCAGAGCGTGTGAAGCAGGAGGGCCCTGCGGTGGCGTCAGGGACCAGGGACA  1027
          |||| || |||||||||||||||| |||||||| ||||||||| ||||| ||||||||
Db   94   GTGTGGTGCCAGGGCGTGTGAAAACAGGACACCTTGCGGTGGTGGTACAGGACAAAGCCA  37
```

RESULT 12

AI558213

```
LOCUS       AI558213            377 bp    mRNA     linear     EST 24-MAR-1999
DEFINITION  vj37a07.y1 Stratagene mouse diaphragm (#937303) Mus musculus cDNA
            clone IMAGE:91188 5' similar to TR:Q63035 Q63035 VASOPRESSIN
            RECEPTOR. ; mRNA sequence.
ACCESSION   AI558213
VERSION     AI558213.1  GI:4508451
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 377)
AUTHORS     Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,T., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
TITLE       The WashU-NCI Mouse EST Project 1999
JOURNAL     Unpublished
COMMENT     Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:536108
            This read is a RESEQUENCE of a previously sequenced mouse clone
            This read has been verified (found to hit its original self in the
```

correct orientation)
Possible reversed clone: similarity on wrong strand
Seq primer: -40RP from Gibco
High quality sequence stop: 369.

FEATURES
source
Location/Qualifiers
1..377
/organism="Mus musculus"
/mol_type="mRNA"
/db_xref="taxon:0090"
/clone="IMAGE:93188"
/tissue_type="diaphragm"
/dev_stage="adult"
/lab_host="SOLR (kanamycin resistant)"
/clone_lib="Stratagene mouse diaphragm (#937303)"
/note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
prepared from diaphragm muscle. Primer: Oligo dT. Average
insert size: 1.5 kb. Uni-ZAP XR Vector; -5' adaptor
sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
CTCGAGTTTTTTTTTTTTTTTT 3'"

BASE COUNT     88 a     114 c     93 g     82 t
ORIGIN

Query Match         7.5%;    Score 200.6; DB 9;   Length 377;
Best Local Similarity   70.9%;   Pred. No. 4e-10;
Matches  266;   Conservative   0;   Mismatches  109;   Indels    0;   Gaps    0;

```
Qy   1558 GCCTCCTCCACAACAGGCTCAGTGAGGCGGGACTGCGTATGCTGCGAGTGAGGGCCTAGCCT 1617
          |||||||||||| ||||||  ||||| ||||||||| ||| | |||||||||||||||
Db      1 GCCTCCTCCAGAGCCGGCCGC CACCAACACAGGCCTGCGTTAC GTGCGAAGGCCTGGCTT   6C

Qy   1618 GGCCGCAGTGCAGGGGTGCAGACGGTCAGGGTACAGTGCCTGACCCCCAGCGAGAGGGCTCC 1677
          ||||   |||| || | |||  |||| ||||| || | |||||||| |||| |||| |||||
Db     61 GGCCCAAGTGCCAGGTGAAGACACTCAGGATGCAGCTCCAGGATCCCAACCTCCAAGAGTGATCA  120

Qy   1678 AGTACCTGGTGGGTATGCTTCGGCCAGAGCCCGCCTGACCACACCCTGGATCTCAGCGGGCT 1737
          ||||| || || ||| || |||||||| ||| | ||||| || | || ||| |||||||||
Db    121 ATTACTTGGTCATCGTGCTCCAGCGAGGCCCAGTCCTGCCAGCCCCTTGGGACCCAGTGGCT  180

Qy   1738 GCCAACTGCCCGCCCCATGGTGACCTACCTGTGCAGTCCGTGCAGCACCAGGGATGCG 1797
          | ||||| || | |||| || | |||||| || |||||| |||||||  |||| |||
Db    181 GTCAGGTGCCTGGCGTCATGTGGAACCTTCCTGTGCAGCCCCCTAAATGCT  240

Qy   1798 GCCTGCAGACCCTCAGTGCTGGCCTCCTGTGGAGCCGAGCGGAGCGGAGCAGTCACTCACAGGAGCTTC 1857
          |||| || || ||| | ||| |||| || || ||| ||| |||||||| |||||  ||||  |||
Db    241 CCCTAAAGACCCTCAGTGCTCTGACCCTCTGTGGAGCTGAGTGAGGAACTCACTTGAGGACTTC  300

Qy   1858 AGGCTGAAGAGAGCAAAGCCGGATCTGGGTCATCACACACCAGCGGCTGAGCGCCACC 1917
          |||||||||||||| ||| ||| |||| |||||   ||||||  |||| ||||||
Db    301 AAGCTGTGAAGCATCAAAGCCAGATCTGTCCATCATATATTCAAAATAAGGCCACACCTC  360

Qy   1918 CACAACCCCCAGG 1932
          ||||| | |||||
Db    361 ACCGGCCTCGGAGG  375
```

RESULT 13
LOCUS       BY281505              431 bp    mRNA     linear   EST 11-DEC-2002
DEFINITION  BY281505 RIKEN full-length enriched, visual cortex Mus musculus
cDNA clone K430354M18 5', mRNA sequence.
ACCESSION   BY281505
VERSION     BY281505.1  GI:26471842
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1  (bases 1 to 431)
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
Nikaido,I., Osato,N., Saito,N., Suzuki,H., Yamanaka,I., Kiyosawa,H.
, Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,

Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
, Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlested,C., Wang,Y.,
Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
, Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.

TITLE      Analysis of the mouse transcriptome based on functional annotation
of 60,770 full-length cDNAs
JOURNAL    Nature 420, 563-573 (2002)
MEDLINE    22354683
PUBMED     12466851

COMMENT    Contact: Yoshihide Hayashizaki
Laboratory for Genome Exploration Research Group, RIKEN Genomic
Sciences Center(GSC), Yokohama Institute
The Institute of Physical and Chemical Research (RIKEN)
1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
Tel: 81-45-503-9222
Fax: 81-45-503-9216
Email: genome-res@gsc.riken.go.jp,
URL:http://genome.gsc.riken.go.jp/
Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
, Murata,M., Nakamura,H., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
, Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
,K., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
Submission.
Computational Analysis of Full-Length Mouse cDNAs Compared with
Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
Normalization and subtraction of cap-trapper-selected cDNAs to
prepare full-length cDNA libraries for rapid discovery of new
genes. Genome Res. 10 (10), 1617-1630 (2000)
RIKEN integrated sequence analysis (RISA) system-384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues.
Tissues were provided by Michela Fagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.

FEATURES             Location/Qualifiers
source               1..431
/organism="Mus musculus"
/mol_type="mRNA"
/strain="C57BL/6J"
/db_xref="taxon:10090"
/clone="K430354M18"
/tissue_type="visual cortex"
/clone_lib="RIKEN full-length enriched, visual cortex"

BASE COUNT     109 a     113 c     114 g     95 t

ORIGIN

Query Match      7.4%;  Score 199.4;  DB 13;  Length 43;
Best Local Similarity 69.5%;  Pred. No. 4.7e-10;
Matches 3C;  Conservative  0;  Mismatches 126;  Indels  6;  Gaps  2;

Qy    522 AGGTCAGTG?ACCTGCT??CATCACCAGGTTC?GAGC?GGC?CCGTAGCCGACGGG 581
          ||||| ||| ||||||  ||||||||||||||||||||||||| ||| |||||||||
Db      2 ACTTCTG?G?ACCTGCTCTT?ATCACCA?CATGCTGA---AGTC?GCAGGCACCAATGGA 58

Qy    582 CCCGGTTGCAGGGCGGACCTGGCAATCTGTGCCGGCCGGAGGGGTCCTCG-- 639
          | ||||| ||| ||||| ||||||||||||||||||||||| |||| |||||||
Db     59 CCCCGGTTGCAGGGAGAGCTGAGAACGCTGTGTGTCGCCTGCCCAGGAGGCATCCTGAT 116

Qy    640 -GACGGCAGGGCGGCAGTTGCCGAGGAGGAAGCGAGCAACTCGGAGGCACCTCGTGGCTCCAAA 696
          |||||||||||||||||||||||||||||||||||||||||||||
Db    119 CATCATAAAGACACAATTCTCAGAAGAGGAGGACCTGGAGAAATTGAAGCTCGGGGTTCCCAA 179

Qy    699 GTGCGAGACCGCTGTT?CTCAGCAAAAAGGAGCTGCCGGGCGTGCCGGAGACAGAGGTCACC 755
          ||||| ||| ||||||||||||||||||||||||||||||||||||||||
Db    179 GT?CAGACAATA?TT?CTCAACAAGAAAGATACCAGGAGTGCTAAAAACTGAGG?CACC 238

Qy    755 TACCAGTTCA?CGACCAGAGCTTCCAGAGTT?CTCG?GGCACTGC?GCTCCTACTGC?GGAG 818
          ||||||||||| || |||||||||||||| || |||||||| || |||||
Db    239 ?ACCAGT?CATTGACCAGAGCTTCCAGGAGT?CTG?GCTGCA?GTCATGATCTGACTACTAGAA 298

Qy    819 GACGCGGGGTGCCCAGGACCGGCCG?GGGGCGGTGGACACTCC?GCGTGGGGACGCC 878
          ||||||| || |||| ||||||||||||||| |||||||||||||
Db    239 GCTGAGGAACTCGGGGACGCCCGCAGGCGCGGGTGTGCAGAAGCTCC?GAACTCGACGCG 358

Qy    979 CAGCCGCACAGCCACTTGGTGTCTCACCACGCGCT?CCTCT?CGGAC?GCTGAGGCGGGAG 938
          |||||||||||| |||||||||||||||||||||||||
Db    359 GAGCTACGTGGTCATCT?GCACTGACCACCGATTCCTC?TGGCCTGC?AAATACAGAG 4:8

Qy    939 CGGATGCGCGACA 951
          ||||| ||||||
Db    419 GGGCT?CGTGACA 431

RESULT 14
AA541951
LOCUS        AA541951        532 bp   mRNA      linear    EST 30-JUL-1997
DEFINITION   vj37a07.r1 Stratagene mouse diaphragm (#937300) Mus musculus cDNA
             clone IMAGE:931188 5' similar to TR:G202806 G202806 VASOPRESSIN
             RECEPTOR. ; mRNA sequence.
ACCESSION    AA541951
VERSION      AA541951.1  GI:2288585
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 532)
  AUTHORS    Marra,M., Hillier,L., Allen,M., Bowles,M., Dietrich,N., Dubuque,T.,
             Geisel,S., Kucaba,T., Lacy,M., Le,M., Martin,J., Morris,M.,
             Schellenberg,K., Steptoe,M., Tan,F., Underwood,K., Moore,B.,
             Theis,B., Wylie,T., Lennon,G., Soares,B., Wilson,R. and
             Waterston,R
  TITLE      The WashU-HHMI Mouse EST Project
  JOURNAL    Unpublished
COMMENT      Contact: Marra M/Mouse EST Project
             WashU-HHMI Mouse EST Project
             Washington University School of Medicine?
             4444 Forest Park Parkway, Box 8501, St. Louis, MO 63:08
             Tel: 314 286 1800
             Fax: 314 286 1810
             Email: mouseest@watson.wustl.edu
             This clone is available royalty-free through :LNL ; contact the
             IMAGE Consortium (info@image.llnl.gov) for further information.
             MGI:536108
             Seq primer: -28m13 revi ET from Amersham
             High quality sequence stop: 312.
FEATURES     Location/Qualifiers

source       1..532
             /organism="Mus musculus"
             /mol_type="mRNA"
             /db_xref="taxon:10090"
             /clone="MAGE:931188"
             /tissue_type="diaphragm"
             /dev_stage="adult"
             /lab_host="SOLR (kanamycin resistant)"
             /clone_lib="Stratagene mouse diaphragm (#937303)"
             /note="Organ: diaphragm; Vector: pBluescript SK-; Site_1:
             EcoRI; Site_2: XhoI; Cloned unidirectionally from mRNA
             prepared from diaphragm muscle. Primer: Oligo dT. Average
             insert size: 1.5 kb. Uni-ZAP XR Vector: -5' adaptor
             sequence: 5' GAATTCGGCACGAG 3' -3' adaptor sequence: 5'
             CTCCAGTTTT?ATTTTTTTTTT 3'."

BASE COUNT      139 a   148 c   129 g   116 t
ORIGIN

Query Match      7.4%;  Score 198.8;  DB 9;  Length 532;
Best Local Similarity 67.9%;  Pred. No. 4.6e-10;
Matches 323;  Conservative  0;  Mismatches 147;  Indels  6;  Gaps  3;

Qy    558 GCCTCCTCCAACAGGCTCAGTGAGCGGGACTGGCGTA?CTGAGTGAGGGCCTAGCCT 1617
          ||||| ||||||||||||||||||||||||| |||||||||||||||||
Db      1 GCCTCCTCCAGAGCCGGCTCACCAACAACAGGCTGGTTTACTGTGCGAAG--CCTGGCT 58

Qy   1618 GGGCCAGTGCAGGGTGCAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCC 1677
          || ||||||||| ||||| |||||||||||||||||||||||||||
Db     59 TGGCCAAGTGCCAGTGAAGACACTCAGGATGCAGGGTGCCTGCAGGGTGATCA 118

Qy   1678 AGTACCTGGTGGTATCT?CGGCAGGCCCCGCCCTGGATCTCAGCGGGCT 1737
          |||||||||||||||||||||
Db    119 ATTACTTGGTCATCGTGCT?CAGCAGAGCCCAGTCCTGACCCTCAGTCGGCT 178

Qy   1738 GCCAACTGCGCGCCCCCATGGTGACCTACCTGTGTGCAGTGAGCTCTGTGGAGCTGGAGCGAGCGAGTCACTACGGATGCG 1797
          ||||||| ||||||||||||||||||||||||||||||
Db    179 GT?CAGCTGCCTGGCGTCA?TGTGAGCGTCA?TGTGAGCTGGAGCTGAGTGAGGCCACACCTAAATGC? 238

Qy   1798 GCCTGCAGACCCTCAGTCTGCCTCTGTGGAGCTTGTGGAGTCGAGGGATGGGGAGGAG 1857
          || ||||||||||||||||||||||
Db    239 CCCTAAAGACCCTCAGTCAGT?CGACCC?CTGTGAGGCTGAGTGAGAACTCACTGAGGGGATCTTC 298

Qy   1858 AGGCTGTGAAGAGAGCAAAGCCGGATCTGGTCATCACGGCTGGCGGCCACC 1917
          ||||||||||||||||||||||||||||||
Db    299 AAGCTGTGAAGACATCAAAGCCAGA?CTGTCCA?CATATATTCAAAATAAGGCACACCTC 358

Qy   1918 CACAACCTCCAAGGAACTCATCTCGACCTTCGAGGCTTCGGTGGCTGGCCAGAGCAGGGTGG 1977
          |||||||||||||||||||||||||||||||||||||
Db    359 ACGGCG?TCGGAAGG-GTGAAGCAGTGTCT?TC?GAAGCTCCGGAGTCCAGAACACGATGA 417

Qy   1978 AAGACCC---TAGTCAAAGTCCCTGTGGAGAACCGCCCATTCCAAGGGCAGGAG 2030
          ||| ||||||||||||||||||||
Db    418 AAGA?GCTGTTAGCCAAAGTCCTTCTAATATGAACTTCCCATTCACGAGCAAGAAG 473

RESULT 15
BX092006
LOCUS        BX092006        495 bp   mRNA      linear    EST 23-JAN-2003
DEFINITION   BX092006 NCI_CGAP_Kid11 Homo sapiens cDNA clone IMAGp998M175911 ;
             IMAGE:2380600, mRNA sequence.
ACCESSION    BX092006
VERSION      BX092006.1  GI:27825094
KEYWORDS     EST.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 495)
  AUTHORS    Ebert,L., Heil,O., Hennig,S., Neubert,P., Partsch,E., Peters,M.,
             Radelof,U., Schneider,D. and Korn,B.
  TITLE      Human UnigeneSet - RZPD3
  JOURNAL    Unpublished
COMMENT      Contact: Ina Rolfs

RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Im Neuenheimer Feld 580, D-69120 Heidelberg, Germany
RZPD; IMAGp998M175911.
RZPDLIB; I.M.A.G.E. cDNA Clone Collection;
Human UnigeneSet - RZPD3 (RZPDLIB No.972)
http://www.rzpd.de/CloneCards/cgi-
bin/showlib.pl?cgi/response?libNo=972 Contact: Ina Rolfs
RZPD Deutsches Ressourcenzentrum fuer Genomforschung GmbH
Heubnerweg 6, D-14059 Berlin, Germany
Tel: +49 30 32639 101
Fax: +49 30 32639 111
www.rzpd.de
This clone is available royalty-free from RZPD;
contact RZPD (clone@rzpd.de) for further information. Seq primer:
M13r, Primer sequence: TTTCACACAGAAACAGCTATGAC.

FEATURES             Location/Qualifiers
     source          1..495
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGp998M175911 ; IMAGE:2380600"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Kid11"
                     /note="Organ: kidney; Vector: pT7T3D-Pac (Pharmacia) with
                     a modified polylinker; Site 1: Not I; Site 2: Eco RI;
                     plasmid DNA from the normalized library NCI_CGAP Kid3 was
                     prepared, and ss circles were made in vitro. Following HAP
                     purification, this DNA was used as tracer in a subtractive
                     hybridization reaction. The driver was PCR-amplified cDNAs
                     from a pool of 5,000 clones made from the same library
                     (cloneIDs 1322376-1323911, 1456007-1456775, and
                     1500552-1502885). Subtraction by Bento Soares and M.
                     Fatima Bonaldo."

BASE COUNT      109 a      141 c      138 g      107 t
ORIGIN

Query Match        6.4%;  Score 171;  DB 13;  Length: 495;
Best Local Similarity   100.0%;  Pred. No. 1.9e-07;
Matches 171;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy   2480 GAGACTAGGGCAGGTGAGACAGCGCAGAACCCCGTGCTGGTGGAAGCATGACCACA 2539
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GAGACTAGGGCAGGTGAGACAGCGCAGAACCCCGTGCTGGTGGAAGCATGACCACA 60

Qy   2540 TGGTGGGTGAGCAGCCCCATGCACTGACGGTAAATCCCGTGTGGACTCATTCTGTTG 2599
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 TGGTGGGTGAGCAGCCCCATGCACTGACGGTAAATCCCGTGTGGACTCATTCTGTTG 120

Qy   2600 GTTTCTATTACACCTGCCAGGCGTGGTACAATACAGGTGCGTGCTGCTCACAA 2650
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GTTTCTATTACACCTGCCAGGCGTGGTACAATACAGGTGCGTGCTGCTCACAA 171

Search completed: October  3, 2003, 13:31:35
Job time : 5527 secs

CM nucleic - nucleic search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on: October 3, 2003, 05:39:00 ; Search time 982: Seconds
(without alignments)
1:201.098 Million cell updates/sec

Title: US-10-029-347-1
Perfect score: 2689
Sequence: 2689
1 cggacgcgtgggcgcgcagc.........aaaaaaaaaaaaaaaaaaa 2689

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 288711 seqs, 2045481386 residues
Total number of hits satisfying chosen parameters: 5777422

Minimum DB seq length: 0
Maximum DB seq length: 200000000
Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : GenEmbl:*
1: gb_ba:*
2: gb_htg:*
3: gb_in:*
4: gb_om:*
5: gb_ov:*
6: gb_pat:*
7: gb_ph:*
8: gb_pl:*
9: gb_pr:*
10: gb_ro:*
11: gb_sts:*
12: gb_sy:*
13: gb_un:*
14: gb_vi:*
15: em_ba:*
16: em_fun:*
17: em_hum:*
18: em_in:*
19: em_mu:*
20: em_om:*
21: em_or:*
22: em_ov:*
23: em_pat:*
24: em_ph:*
25: em_pl:*
26: em_ro:*
27: em_sts:*
28: em_un:*
29: em_vi:*
30: em_htg_hum:*
31: em_htg_inv:*
32: em_htg_other:*
33: em_htg_mus:*
34: em_htg_pln:*
35: em_htg_rod:*
36: em_htg_man:*
37: em_htg_vrt:*
38: em_sy:*
39: em_htgo_hum:*
40: em_htgo_mus:*
41: em_htgo_other:*

score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 2689 | 100.0 | 2689 | 6 | AX663294 | AX663294 Sequence |
| 2 | 2626 | 97.7 | 3365 | 6 | AX168123 | AX168123 Sequence |
| 3 | 2605 | 96.9 | 2763 | 6 | AX663298 | AX663298 Sequence |
| 4 | 2020.2 | 75.1 | 2054 | 6 | AX663299 | AX663299 Sequence |
| 5 | 1925 | 71.6 | 2679 | 9 | AF479748 | AF479748 Homo sapi |
| 6 | 1925 | 71.6 | 2679 | 9 | AY154461 | AY154461 Homo sapi |
| 7 | 1571.2 | 58.4 | 5553 | 6 | AX088215 | AX088215 Sequence |
| 8 | 1571.2 | 58.4 | 5556 | 6 | AX459875 | AX459875 Sequence |
| 9 | 1571.2 | 58.4 | 6461 | 6 | AX088208 | AX088208 Sequence |
| 10 | 1369 | 50.9 | 95038 | 9 | AF015416 | AF015416 Homo sapi |
| 11 | 1365.8 | 50.8 | 143835 | 9 | AC136475 | AC136475 Homo sapi |
| 12 | 1365.8 | 50.8 | 156819 | 2 | AP006283 | AP006283 Homo sapi |
| 13 | 944.8 | 33.9 | 3454 | 10 | BC031139 | BC031139 Mus muscu |
| 14 | 911.2 | 33.9 | 2354 | 10 | BC013519 | BC013519 Mus muscu |
| 15 | 838.4 | 31.2 | 175416 | 2 | AP006286 | AP006286 Homo sapi |
| 16 | 838.4 | 31.2 | 187160 | 9 | AC138230 | AC138230 Homo sapi |
| 17 | 823.2 | 30.6 | 2257 | 10 | RATAIIR | M85183 Rattus norv |
| 18 | 764.2 | 28.4 | 238533 | 2 | AC132968 | AC132968 Rattus no |
| c 19 | 724 | 26.9 | 251206 | 2 | AC107815 | AC107815 Mus muscu |
| c 20 | 648.4 | 24.1 | 61320 | 2 | AC135992 | AC135992 Homo sapi |
| 21 | 514.4 | 19.1 | 61320 | 2 | AC135992 | AC135992 Homo sapi |
| 22 | 427.6 | 15.9 | 678 | 6 | AR270448 | AR270448 Sequence |
| 23 | 385.2 | 14.3 | 699 | 9 | HSA327212 | AJ327212 Homo sapi |
| 24 | 381.4 | 14.2 | 780 | 9 | HSA331725 | AJ331725 Homo sapi |
| 25 | 362.4 | 13.5 | 765 | 9 | HSA329750 | AJ329750 Homo sapi |
| 26 | 278.6 | 10.4 | 3466 | 9 | AK095460 | AK095460 Homo sapi |
| 27 | 277 | 10.3 | 3507 | 9 | BC028069 | BC028069 Homo sapi |
| 28 | 277 | 10.3 | 3827 | 9 | AY095146 | AY095146 Homo sapi |
| 29 | 277 | 10.3 | 3827 | 9 | AY154467 | AY154467 Sequence |
| 30 | 273.4 | 10.2 | 3108 | 6 | AX417214 | AX417214 Sequence |
| 31 | 273.4 | 10.2 | 3300 | 6 | AX459869 | AX459869 Sequence |
| 32 | 273.4 | 10.2 | 3395 | 9 | AY116206 | AY116206 Homo sapi |
| 33 | 273.4 | 10.2 | 3563 | 9 | AY116205 | AY116205 Homo sapi |
| 34 | 273.4 | 10.2 | 3731 | 9 | AY116204 | AY116204 Homo sapi |
| 35 | 271.4 | 10.1 | 3221 | 9 | AY116207 | AY116207 Homo sapi |
| 36 | 258 | 9.6 | 314 | 6 | AX663300 | AX663300 Sequence |
| c 37 | 240.6 | 8.9 | 147330 | 6 | AC008753 | AC008753 Homo sapi |
| 38 | 236.6 | 8.8 | 257 | 6 | AR269645 | AR269645 Sequence |
| c 39 | 229.6 | 8.5 | 229028 | 2 | AC123316 | AC123316 Rattus no |
| c 40 | 229.6 | 8.5 | 229793 | 2 | AC098184 | AC098184 Rattus nc |
| 41 | 228 | 8.5 | 1958 | 9 | AY051114 | AYC5111233 AY051114 Homo sapi |
| 42 | 228 | 8.5 | 2524 | 6 | AX417249 | AX417249 Sequence |
| 43 | 228 | 8.5 | 2528 | 9 | AF418985 | AF418985 Homo sapi |
| 44 | 228 | 8.5 | 2546 | 9 | AK027194 | AK027194 Homo sapi |
| 45 | 228 | 8.5 | 3237 | 9 | AY092033 | AY092033 Homo sapi |

ALIGNMENTS

RESULT 1
AX663294
LOCUS       AX663294                2689 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 1 from Patent WO02061086.
ACCESSION   AX663294
VERSION     AX663294.1  GI:29163660
KEYWORDS
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Feder,J., Ramanathan,C. and Mintier,G.
TITLE       Human leucine-rich repeat containing protein, expressed
            predominately in small intestine, HLRRSI1

Pred. No. is the number of results predicted by chance to have a

JOURNAL    Patent: WO 02061086-A 1 08-AUG-2002;
           Bristol-Myers Squibb Company (US)
FEATURES   Location/Qualifiers
source     1..2689
           /organism="Homo sapiens"
           /mol_type="genomic DNA"
           /db_xref="taxon:9606"
CDS        75..1952
           /note="unnamed protein product"
           /codon_start=1
           /protein_id="CAD80012.1"
           /db_xref="GI:29163661"
           /translation="M:AQPQRLLFILDGADELPALGGPEAAPCTPFEAASGARVLGG
           LLSKALLPTALLLVTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFKFFRRERRA
           ERAYRFVKENETLFALCEVFPVCWIVCTVLRQQLEIGRDLSRTSKTTSVYLLFITSV
           LSSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELECLELRGSKVCTLFLSKKE
           LPGVLTEVTYQF:DCSFQEFLAALSYLLEDGGVPRTAAGGVGTLLRGPACPHSHLVL
           TTRPLFLGLLSAERMRDIERHFGCYVSERVKQEALRVVQGQSQGCPAVAPEVTEGAKGL
           EDTEEPEEEGEEFNYPLELLYC:YETQEDAFVRQALCRFPE:ALQRVRFCRMCVAV
           LSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKPLQASLGGGSSCGT:KQLPASLLH
           PLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPARHTELGLHHRISEAGLRM
           LSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQSPALTTLDLSCCQLPAPMVTYLC
           AVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDLVITHPALLDGHPQPPKELIGT
           F"

BASE COUNT    534 a    861 c    844 g    450 t
ORIGIN

Query Match        100.3%;    Score 2689;    DB 6;    Length 2689;
Best Local Similarity    100.3%;    Pred. No. 0;
Matches 2689;    Conservative    0;    Mismatches    0;    Indels    0;    Gaps    0;

SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS     Burford,N., Baughn,M.R., Au-Young,J., Yang,J., Lu,D.A. and Reddy,R.
TITLE       G-protein coupled receptors
JOURNAL     Patent: WO 0142288-A 41 14-JUN-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..3365
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 1499408CB1"
BASE COUNT      612 a    1106 c    1114 g     533 t
ORIGIN

Query Match       97.7%;   Score 2626;   DB 6;   Length 3365;
Best Local Similarity   99.8%;   Pred. No. 0;
Matches 2629;   Conservative   0;   Mismatches   0;   Indels   5;   Gaps   0;

```
Qy      7 CGTGGGCGCCGCAGCCTGCTGACCTGATCCTGGACCAGTGCCCGCCGCCGGCCCGCCCGG   66
          |||||  ||||| || ||||||||||||||||||||||||||||||||||  || |||||
Db    731 CGGGCACGCGCCTGGCTGACCTGATCCTGGACCAGTGCCCGCCGCCGGCCGACGGCCCGG  790

Qy     67 TGCCGCAGATGCTGGCCCAGCCGCTGCTTCATCCTGGACGGCGCCGGACGGAGC        126
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 TGCCGCAGATGCTGGCCCAGCCGCTGCTTCATCCTGGACGGCGCCGGACGGAGC        850

Qy    127 TGCCGGCGCTGGGGGGCCGCGCCCGAGGCGCCCCTTGCAGACACCCTTGAGGCGCGAGCG  186
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 TGCCGGCGCTGGGGGGCCGCGCCCGAGGCGCCCCTTGCAGACACCCTTGAGGCGCGAGCG  910

Qy    187 GCGGCGGGTGCTAGGCGCGCGCCCGGAGGCGCTGAGTAAGGCGCCTCTGCTGG        246
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCGGCGGGTGCTAGGCGCGCGCCCGGAGGCGCTGAGCAAGGCGCTGCTCCTGCTGG      970

Qy    247 TGACCACCGGCGCGCCGGCTTCTCCGAGCGCGCCCGGGGAGGCGCTGTGTCCGTCCAGTGCG  306
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 TGACCACCGGCGCGCCGGCTTCTCCGAGCGCGCCGCTGTGTGTCCCGCCAGTGCG  1030

Qy    307 CCGAGGTGCGCGGCTTCTCCGAGCGCGCCGCAGACAAGGACAAGGAGAAGAGTATTTCTACAAGTTCTTCCGGG  366
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031 CCGAGGTGCGCGGCTTCTCCGAGCGCGCCGCAGACAAGGACAAGGAGAAGAGTATTTCTACAAGTTCTTCCGGG  1090

Qy    367 ATGAGAGAGGGCGCAGGGCGCCTAGGCCTATGGCCCAGGGCCCACAGGCGCCGAGAACGAGACGAGAC         426
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091 ATGAGAGAGGGCGCAGGGCGCCTAGGCCTATGGCCCAGGGCCCACAGGCGCCGAGAACGAGACGAGAC       1150

Qy    427 TGTGTTCGTGCCCTTCGTGTGCGGATCGTGTGCACCGTCCAAGACCGCTCAGTGGCGCGCCAGCAGCTGGAGC  486
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151 TGTGTTCGTGCCCTTCGTGTGCGGATCGTGTGCACCGTCCAAGACCGCTCAGTGGCGCGCCAGCAGCTGGAGC  1210

Qy    487 TCGGTGGGACCTGTCGGCGGCACCTCCAAGACCAAGCCCGGTTGCAGGGCGACCTGGCGA     546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211 TCGGTGGGACCTGTCGGCGGCACCTCCAAGACCAAGCCCGGTTGCAGGGCGACCTGGCCA   1270

Qy    547 CCAGCGTTCTGAGTCTGGCTCCGGCCCGGAGGGGCGTCCTGGACGGGGCCCGAGCGGCAGTTGCCGAGAAGG  606
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1271 CCAGCGTTCTGAGTCTGGCTCCGGCCCGGAGGGGCGTCCTGGACGGGGCCCGAGCGGCAGTTGCCGAGAAGG  1330

Qy    607 ATCTGTGCCGCCTGGCCCTGGCCCCGAGGGGCGTTGCCGAGAAGG        666
          |||||||||||||||||||||||||||||||||||||||||||||
Db   1331 ATCTGTGCCGCCTGGCCCTGGCCCCGAGGGGCGTTGCCGAGAAGG      1390

Qy    667 AACTGGAGCAACTGGAGCTTCGTGCCTCCAAAGTGCAGACGCGTGTTTCTCAGCAAAAAGG  726
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1391 AACTGGAGCAACTGGAGCTTCGTGCCTCCAAAGTGCAGACGCGTGTTTCTCAGCAAAAAGG  1450

Qy    727 AGCTGCCGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG  786
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1451 AGCTGCCGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG  1510
```

```
Qy   1741 AACTGCCCGCCCCATGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGCC  1800
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741 AACTGCCCGCCCCCATGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGCC  1800

Qy   1801 TGCAGACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGGCAGCGAGTCACTACAGAGACC TCAGG  1860
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801 TGCAGACCCTCAGTCTGGCC CGTGGAGCTGAGCGGCAGTCACTACAGAGACC TCAGG  1860

Qy   1861 CTGTGAAGAGAGCAAAGCCGGATCTGTCATCACACACCCAGCGCGTGGACGGCCACCAC  1920
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861 CTGTGAAGAGAGCAAAGCCGGATCTGTCATCACACACCCAGCGCGTGGACGGCCACCAC  1920

Qy   1921 AACCTCCCAAGGAACTCATCTCGACCCTTCTGAGGCTTCGTGGCCAGAGCAGGGTGGAAG  1980
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921 AACCTCCCAAGGAACTCATCTCGACCCTTCTGGTGGCCAGAGCAGGGTGGAAG  1980

Qy   1981 ACCCTAGTCAAAGTCCCTGTGAGAGAACGGCCCATTCCAAGGGCAGGAGGAGATATTGCTC  2040
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1981 ACCCTAGTCAAAGTCCCTGTGAGAGAACGGCCCATTCCAAGGGCAGGAGGAGATATTGCTC  2040

Qy   2041 TCGGCCTTTGGGAAACTTTTGAGCCCAGAACAGCAGGCCGCGAGAGGCCGCGAGGAGCGCCCAGAC  2100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2041 TCGGCCTTTGGGAAACTTTGAGCCCAGAACAGCAGGCATGTGGGAGGAGCGCCCAGAC  2100

Qy   2101 ACGGCACCCTGCCCCGTCCAGACAGGCCCAGGACCTGCCCCTCTCTCCACACCTGGGGT  2160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 ACGGCACCCTGCCCCGTCCAGACAGGCCCAGGACCTGCCCCTCTCTCCACACCTGGGGT  2160

Qy   2161 ACCCCTTCTCCCCAGCCCACACTACTCCCCACCTTCCTCTCCTGAGACCCTCCAG      2220
          ||||||||||||||||||||  ||||||||||||||||||||||||||||||||
Db   2161 ACCCCTTCTCTCCCCAGCCCCA C CACTACTCCCCACCTTCCTC TCCTGAGACCCTCCAG  2220

Qy   2221 CCATTCCCTTGAAAACACCCCCGACCCCCAAGCCACAATAATGACAGGAGAGCTCCAA  2280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 CCATTCCCTTGAAAACACCCCCGACCCCCAAGCCACAATAATGACAGGAGAGCTCCAA  2280

Qy   2281 TTAACTAAGCACCTACCTGGCGGCAGAATAACCCTCACTGCCTGATCCCCATTCTGCAGT  2340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 TTAACTAAGCACCTACCTGGCGGCAGAATAACCCTCACTGCCTGATCCCCATTCTGCAGT  2340

Qy   2341 GTGGCCCAACAGCCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGTCC  2400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 GTGGCCCAACAGCCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGTCC  2400

Qy   2401 CTCCCTGTTAGGAATGAGACCATCCCTGAGGCTATGGCCCAGGGCCCACAGGCGCGGAACCCCCGTGCTG  2460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 CTCCCTGTTAGGAATGAGACCATCCCTGAGGCTATGGCCCAGGGCCCACAGGCGCGGAACCCCCGTGCTG  2460

Qy   2461 TCTGAGATCTTTGGGAAGGGAGGACTAGGGCAGGTGAGCAGCCCCATGCACTGAC GGTAAATTCCC  2520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 TCTGAGATCTTTGGGAAGGGAGGAGACTAGGGCAGGTGAGCAGCCCCATGCACTGACGGTAAATTCCC  2520

Qy   2521 GGTGGGAAGCATGACCACCATGGGGGTGAGCAGCCCCCATGCACTGACGGTAAATTCCCC  2580
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 GGTGGGAAGCATGACCACCATGGTGGGTGAGCAGCCCCCATGCACTGACGGTAAATTCCCC  2580

Qy   2581 TGTGGACTCATTTCTGTGGGTTTCTATTACACCTGGCCAGGCGTCGGTACAATACAGGTCG  2640
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 TGTGGACTCATTTCTGTGGGTTTCTATTACACCTGGCCAGGCGTGGTACAATACAGGTCG  2640

Qy   2641 GTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2689
          |||||||||||||||||||||||||||||||||||||||||||||
Db   2641 GTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2689
```

RESULT 2
AX168123
LOCUS       AX168123     3365 bp     DNA     linear     PAT 03-JUL-2001
DEFINITION  Sequence 41 from Patent WO0142288.
ACCESSION   AX168123
VERSION     AX168123.1     GI:14597401
KEYWORDS    .

RESULT 3
AX663298

LOCUS       AX663298                2763 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 5 from Patent WO02061086.
ACCESSION   AX663298
VERSION     AX663298.1  GI:29163662
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Feder,J., Ramanathan,C. and Mintier,G.
  TITLE     Human leucine-rich repeat containing protein, expressed
            predominantly in small intestine, HLRRSI1
  JOURNAL   Patent: WO 0261086-A 5 08-AUG-2002;
            Bristol-Myers Squibb Company (US)
FEATURES             Location/Qualifiers
     source          1..2763
                     /organism="Homo sapiens"

```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      544 a    890 c    870 g    459 t
ORIGIN

  Query Match             96.9%;  Score 2605;  DB 6;  Length 2763;
  Best Local Similarity   97.3%;  Pred. No. 0;
  Matches 2689;  Conservative    0;  Mismatches    0;  Indels   74;  Gaps    1;

Qy          1 CGGACGCGTGGGCGCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCG 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          1 CGGACGCGTGGGCGCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCG 60

Qy         61 CGCCGGTGCCGCAGATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGG 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         61 CGCCGGTGCCGCAGATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGG 120

Qy        121 ACGAGCTGCCGGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGG 180
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        121 ACGAGCTGCCGGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGG 180

Qy        181 CGAGCGGCGCGCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCCACGGCCCTCC 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        181 CGAGCGGCGCGCGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCCACGGCCCTCC 240

Qy        241 TGCTGGTGACCACGCGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGC 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        241 TGCTGGTGACCACGCGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGC 300

Qy        301 AGTGCGCCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCT 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        301 AGTGCGCCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCT 360

Qy        361 TCCGGGATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGT 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        361 TCCGGGATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGT 420

Qy        421 TCGCGCTGTGCTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGC 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        421 TCGCGCTGTGCTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGC 480

Qy        481 TGGAGCTCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTT 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        481 TGGAGCTCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTT 540

Qy        541 TCATCACCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACC 600
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        541 TCATCACCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACC 600

Qy        601 TGCGCAATCTGTGCCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGCGCAGTTTGCCG 660
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        601 TGCGCAATCTGTGCCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGCGCAGTTTGCCG 660

Qy        661 AGAAGGAACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCA 720
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        661 AGAAGGAACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCA 720

Qy        721 AAAAGGAGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCT 780
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        721 AAAAGGAGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCT 780

Qy        781 TCCAGGAGTTCCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACCG 840
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        781 TCCAGGAGTTCCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACCG 840

Qy        841 CGGCTGGCGGCGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTTGGTGC 900
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        841 CGGCTGGCGGCGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTTGGTGC 900

Qy        901 TCACCACGCGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGCC 960
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        901 TCACCACGCGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGCC 960

Qy        961 ACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGAC 1020
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        961 ACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGAC 1020

Qy       1021 AGGGACAGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGG 1080
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1021 AGGGACAGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGG 1080

Qy       1081 ACACCGAAGAGCCAGAGGAGGAGGAGGAGGGAGAGGAGCCCAACTACCCACTGGAGTTGC 1140
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1081 ACACCGAAGAGCCAGAGGAGGAGGAGGAGGGAGAGGAGCCCAACTACCCACTGGAGTTGC 1140

Qy       1141 TGTACTGCCTGTACGAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCC 1200
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1141 TGTACTGCCTGTACGAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCC 1200

Qy       1201 CGGAGCTGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACT 1260
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1201 CGGAGCTGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACT 1260

Qy       1261 GCGTGAGGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTG 1320
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1261 GCGTGAGGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTG 1320

Qy       1321 CGCAGGAGAAGAAGAAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCA 1380
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1321 CGCAGGAGAAGAAGAAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCA 1380

Qy       1381 GTTCTCAAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAA 1440
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1381 GTTCTCAAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAA 1440

Qy       1441 TGACTGACCCACTGTGCCATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACG 1500
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1441 TGACTGACCCACTGTGCCATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACG 1500

Qy       1501 CGGTCTGCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCCGCACTGACGGAGCTGGGCC 1560
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1501 CGGTCTGCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCCGCACTGACGGAGCTGGGCC 1560

Qy       1561 TCCTCCACAACAGGCTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGC 1620
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1561 TCCTCCACAACAGGCTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGC 1620

Qy       1621 CGCAGTGCAGGGTGCAGACGGT-------------------------------------- 1642
              ||||||||||||||||||||||
Db       1621 CGCAGTGCAGGGTGCAGACGGTCAGGTGAGGCCTGGCCTGGGAGGGACCGTGGGATGCCC 1680

Qy       1643 ------------------------------------CAGGGTACAGCTGCCTGACCCCA 1666
                                                  |||||||||||||||||||||||
Db       1681 CCGCCACCCCAGCAGCTCCTGAGGTCGGCCCTCCCACAGGGTACAGCTGCCTGACCCCA 1740

Qy       1667 GCGAGGGCTCCAGTACCTGGTGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCCTGGA 1726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1741 GCGAGGGCTCCAGTACCTGGTGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCCTGGA 1800

Qy       1727 TCTCAGCGGCTGCCAACTGCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCA 1786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1801 TCTCAGCGGCTGCCAACTGCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCA 1860

Qy       1787 CCAGGGATGCGGCCTGCAGACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACT 1846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1861 CCAGGGATGCGGCCTGCAGACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACT 1920

Qy       1847 ACAGGAGCTTCAGGCTGTGAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCT 1906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1921 ACAGGAGCTTCAGGCTGTGAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCT 1980

Qy       1907 GGACGGCCACCCACAACCTCCCAAGGAACTCATCTCGACCTTCTGAGGCTCTGGTGGCCA 1966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1981 GGACGGCCACCCACAACCTCCCAAGGAACTCATCTCGACCTTCTGAGGCTCTGGTGGCCA 2040

Qy       1967 GAGCAGGGTGGAAGACCCTAGTCAAAGTCCCTGTGGAGAGAACGGCCCATTCCAAGGGCA 2026
```

BASE COUNT     473 a     631 c     607 g     343 t
ORIGIN

Query Match           75.1%;   Score 2020.2;   DB 6;   Length 2054;
Best Local Similarity 99.9%;   Pred. No. 3.4e-306;
Matches 2022;   Mismatches 0;   Conservative 0;   Indels 3;   Gaps 0;
```

```
                                                              473 a  631 c  607 g  343 t
                                                              /db_xref="taxon:9606"

Qy    665 GGAACTGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCGTGTTTCTCAGCAAAAA 724
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 GGAACTGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCGTGTTTCTCAGCAAAAA 60

Qy    725 GGAGCTGCCGGGCGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACCAGAGCTTCCA 784
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61 GGAGCTGCCGGGCGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACCAGAGCTTCCA 120

Qy    785 GGAGTTCCTGCGGGCACTGCTCTACCTGCTGGAGGACACGGGCGGGTGCCAGGACCGCGGC 844
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121 GGAGTTCCTGCGGGCACTGCTCTACCTGCTGGAGGACACGGGCGGGTGCCAGGACCGCGGC 180

Qy    845 TGGCGGGTTGGGACACTTCCTGCGTGGGGACGCCAGCCGCACAGCCACTTGGTGCTCAC 904
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181 TGGCGGGTTGGGACACTTCCTGCGTGGGGACGCCAGCCGCACAGCCACTTGGTGCTCAC 240

Qy    905 CACGCGCTTCCTTCCGGACTGCTGAGCGCGAGCCGGATGCGCGACATCGAGCGCGCCACTT 964
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241 CACGCGCTTCCTTCCGGACTGCTGAGCGCGAGCCGGATGCGCGACATCGAGCGCGCCACTT 300

Qy    965 CGGCTGCATGGTTTCAGAGCGTGAAGCAGCAGGGCCCTGCGGCGTGGTGCAGGGGACAGGG 1024
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301 CGGCTGCATGGTTTCAGAGCGTGAAGCAGCAGGGCCCTGCGGCGTGGTGCAGGGGACAGGG 360

Qy   1025 ACAGGGCTGCCCCGGAGTGGCACCAGAGGTGACCAGGAGGCTCGAGGACGACTCGAGGACAC 1084
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361 ACAGGGCTGCCCCGGAGTGGCACCAGAGGTGACCAGGAGGCTCGAGGACGACTCGAGGACAC 420

Qy   1085 CGAAGAGCCAGAGGAGGAGGAGGAGGAGGAGAGCCCAACTACCCACTGGAGTTGCTGTA 1144
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421 CGAAGAGCCAGAGGAGGAGGAGGAGGAGGAGAGCCCAACTACCCACTGGAGTTGCTGTA 480

Qy   1145 CTGCCTTACGAGACGGAGCTTTGTGGCGCAAGCCCGGTTCGCGCCGGTTCCCGGA 1204
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481 CTGCCTGTACGAGACGGAGCTTTGTGGCGCAAGCCCGGTTCGCGCCGGTTCCCGGA 540

Qy   1205 GCTGGCGCTCAGCGAGCTGGCGCTTCTGCCGACGGCGAGTGCGTGTTCTGAGCTACTGCGT 1264
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541 GCTGGCGCTCAGCGAGCTGGCGCTTCTGCCGACGGCGAGTGCGTGTTCTGAGCTACTGCGT 600

Qy   1265 GAGGTGCTGCCCCTGCTCGGACAGGCACTGCAGATTGGTTGCTGCGCA 1324
          |||||||||||||||||||||||||||||||||||||||||||||||
Db    601 GAGGTGCTGCCCCTGCTCGGACAGGCACTGCAGATTGGTTGCTGCGCA 660

Qy   1325 GGAGAAGAGAAGAAGAGCCTGGGAAGCGGCTCCAGCCTCAGCCAGCGGGCCAGTTC 1384
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661 GGAGAAGAGAAGAAGAGCCTGGGAAGCGGCTCCAGCCTCAGCCAGCGGGCCAGTTC 720

Qy   1385 TCAAGGCACCAAAAACAACTGCCAGCCTCCTTCTTCATCGCTCTTCAGGCAATGAC 1444
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721 TCAAGGCACCAAAAACAACTGCCAGCCTCCTTCTTCATCGCTCTTCAGGCAATGAC 780

Qy   1445 TGACCCACTGTGCCATCTGAGCAGCCTCAGCGTGTCCACTGCAAACTCCCTGACGCGGT 1504
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781 TGACCCACTGTGCCATCTGAGCAGCCTCAGCGTGTCCACTGCAAACTCCCTGACGCGGT 840

Qy   1505 CTGCCGAGACCTTTCTGAGGCCCTGAGGCCAGCCCCCGCACTGACGGAGCTGGGCCTCCT 1564
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841 CTGCCGAGACCTTTCTGAGGCCCTGAGGCCAGCCCCCGCACTGACGGAGCTGGGCCTCCT 900

Qy   1565 CCCAAACAGGCTCAGTGAGGCGGGGACTGCGTATGCTGAGTGAGGGCCTGGCGCCGCA 1624
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901 CCCAACAGGCTCAGTGAGGCGGGGACTGCGTATGCTGAGTGAGGGCCTGGCGCCGCA 960

Qy   1625 GTGCAGGGGTGCAGACGGTCAGGTACAGCTGCCTGACGTGCCCCCAGGGCCAGTACCT 1684
```

```
Db   2041 GAGCAGGGGTGGAAGACCCTAGTCAAAGTCCCTGTGGAGAACGGCCCATTCCAGGGCA 2130
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   2027 GGAGGGATATTGCTCTCGGCCTTGGGAAACTTTGAGCCGAGAGGCCCAGACAGGCATG 2086
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2101 GGAGGGATATTGCTCTCGGCCTTGGGAAACTTTGAGCCGAGAGGCCCAGACAGGCATG 2160

Qy   2087 TGGGAGGGCCAGACACGGCACCTGCCCGGTCCGCCAGGACAGAGGCCCAGGACCTGCCCCCTC 2146
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2161 TGGGAGGGCCAGACACGGCACCTGCCCGGTCCGCCAGGACAGAGGCCCAGGACCTGCCCCCTC 2220

Qy   2147 TCCACCTGGGGTACCCCTTCTCCCCCACCACTACCCACCCCTTCCCCTC 2206
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2221 TCCACCTGGGGTACCCCTTCTCCCCCACCACTACCCACCCCTTCCTCTC 2280

Qy   2207 CTGAGACCCCAGCCATCCCCTGAAAACACCCCCAGCCACAATAATGAC 2266
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   2281 CTGAGACCCCAGCCATCCCCTGAAAACACCCCCAGCCACAATAATGAC 2340

Qy   2267 AGCGAGAGCTCCAATTAACTAAGCACCTACCTGGCGGCGAGAATAACCCTTCACTGCCGA 2326
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2341 AGCGAGAGCTCCAATTAACTAAGCACCTACCTGGCGGCCAGAATAACCCTTCACTGCCTGA 2400

Qy   2327 TCCCATCTGCAGTGTGGCCCAACAGCCCCAGAACTATGCCCACATAGACTGGAGGTAG 2386
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2401 TCCCATCTGCAGTGTGGCCCAACAGCCCCAGAACTATGCCCACATAGACTGGAGGTAG 2460

Qy   2387 GCAGTTCACCGTCCCCCTTAGGAATGAGACATCCCCGAGGCTATGGCCCAGGCCC 2446
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2461 GCAGTTCACCGTCCCCCTGTTAGGAATGAGACATCCCCGAGGCTATGGCCCAGGCCC 2520

Qy   2447 ACAGGGCGTCCAGTGTCTCAGACTTTGGGAAGGGAGACTAGGGGACTGGAGACAGCGCA 2506
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2521 ACAGGGCGTCCAGTGTCTCAGACTTTGGGAAGGGAGACTAGGGGACTGGAGACAGCCGCA 2580

Qy   2507 GAACCCCGTGCTGGGTGGGAAGCATGACCACCATGGTTCATTACACCTGGCCCAGGCGTGG 2566
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2581 GAACCCCGTGCTGGGTGGGAAGCATGACCACCATGGTTCATTACACCTGGCCCAGGCGTGG 2640

Qy   2567 ACGGTAAATTCCCCTGGACTCATTTCTGTTGGTTTCATTACACCTGGCCCAGGCGTGG 2626
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2641 ACGGTAAATTCCCCTGGACTCATTTCTGTTGGTTTCATTACACCTGGCCCAGGCGTGG 2700

Qy   2627 TACAATACAGGTCGGTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2686
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   2701 TACAATACAGGTCGGTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2760

Qy   2687 AAA 2689
          |||
Db   2761 AAA 2763
```

```
RESULT 4
AX663299
LOCUS       AX663299    2054 bp    DNA    linear    PAT 22-MAR-2003
DEFINITION  Sequence 6 from Patent WO02061086.
ACCESSION   AX663299
VERSION     AX663299.1  GI:29163663
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Feder,J., Ramanathan,C. and Mintier,G.
  TITLE     Human leucine-rich repeat containing protein, expressed
            predominantly in small intestine, HLRRS11
  JOURNAL   Patent: WO 02061086-A 6 08-AUG-2002;
            Bristol-Myers Squibb Company (US)
FEATURES             Location/Qualifiers
     source          1..2054
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
```

```
AF479748
LOCUS       AF479748                2679 bp    mRNA      linear   PRI 12-AUG-2002
DEFINITION  Homo sapiens PYRIN-containing APAF1-like protein 5 mRNA, complete
            cds.
ACCESSION   AF479748
VERSION     AF479748.1  GI:19387135
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2679)
  AUTHORS   Wang,L., Manji,G.A., Grenier,J.M., Al-Garawi,A., Merriam,S.,
            Lora,J.M., Geddes,B.J., Briskin,M., DiStefano,P.S. and Bertin,J.
  TITLE     PYPAF7, a Novel PYRIN-containing Apaf1-like Protein That Regulates
            Activation of NF-kappa B and Caspase-1-dependent Cytokine
            Processing
  JOURNAL   J. Biol. Chem. 277 (33), 29874-29880 (2002)
  PUBMED    12019269
REFERENCE   2  (bases 1 to 2679)
  AUTHORS   Bertin,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (01-FEB-2002) Inflammation, Millennium Pharmaceuticals
            Inc., 45 Sidney Street, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..2679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p15"
     CDS             1..2679
                     /note="PYPAF5; PYRIN-containing nucleotide-binding
                     site/leucine-rich repeat family member"
                     /codon_start=1
                     /product="PYRIN-containing APAF1-like protein 5"
                     /protein_id="AAL87105.1"
                     /db_xref="GI:19387136"
                     /translation="MDQPEAPCSSTGPRLAVARELLLAALELSQEOLKsPRHKLRDV
                     GPDGRSIPWGRLERADAVDLAEQLAQFYGPEPALEVARKTLKRADARDVAAQLQERRL
                     QRLGLSGTLLSVSEYKKKYREHVLQLHARVKERNARSVKITKRFTKLLIAPESAAPE
                     EALGPAEEPEPGRARRSDTHFNRLFRRDEGRRPLTVLQGPAGIGKTVAAKKILYD
                     WAAGKLYQGQVDFAFFMPCGELLERPGTRSLADLILDQCPDRGAPVPQMLAQPQRLLF
                     ILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTA.LLVTRAAAPGRL
                     QGRLCSPQCAEVRGFSKDKKKYFYKFFRDERRAERAYRFVKENETLFALCFVPFVCW
                     IVCTVLRQQLEIGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRIQGDLRNLCRLAR
                     EGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKE.PGVLETEVTYQFIDQSFQEFLAA
                     LSYLLEDGSVPRTAAGSVGTLLRGDAQPHSHiVLTTRFLFGLLSAERMRDIERHFSCV
                     VSERVKOEALRWVQGQGOGCPGVAPEVTEGAKGLEDTEPPEEEBEGGEEPNYPLELLYC
                     LYETQEDAFVRQALCRFPPELALQRVRFCRMDVAVLSYCVRCCPAQALRLiSCRLVRA
                     QEKKKKSLGKRLQASLGGSSSQGTTKQLPASLLHPLFQAMTDPLCHLSSSLTLSHCKL
                     PDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQVTVRVQLPDPQ
                     RGiQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHGGCGLQTLSLASVELSEQ
                     SLQELQAVKRAKPDLvITHPALDGHPQPPKELISTF"
BASE COUNT      458 a    878 c    934 g    409 t
ORIGIN
```

```
Query Match             71.6%;  Score 1925;  DB 9;  Length 2679;
Best Local Similarity   99.6%;  Pred. No. 2.5e-291;
Matches 1941;  Conservative    0;  Mismatches    0;  Indels    5;  Gaps    3;

Qy     7  CGTGGGCCCCAGCCTGGCTACCTGACCTGATCCTGACCAGTGCCCGCGCCGGACCGGACCGGCGGCCGCCGCCGG   66
           ||||| |||||||||||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   731  CGGGCACCGCCAGCCTGGCTGACCTGACCTGATCCTGACCAGTGCCCGCGCCGGACCGGACCGGCGGCCGCCGCCGG  790

Qy    67  TGCCCGCAGATGCTGCCCCAGCCGGCCGCCGCCGCCGCCGCCGCCGCCGACGAGC  126
           ||||||||||||||||||||||||||||
Db   791  TGCCCGCAGATGCTGCCCCAGCCGGCCGCTGCCGCCGCTGCCGCCGGGCGGGACGAGC  850

Qy   127  TGCCGGCGTGGGGGGCCCCGAGGCCGG  186
           |||| ||||||||||||||||||||||
Db   851  TGCCGGCGTGGGGGGCCCCGAGGCCCTTGCACAGACCCCTTGCACAGACCCCTTGAGGGCGGAGCG  910
```

```
Db   961  GTGCAGGGGCGCAGCGGTCAGGGTACAGCTGCCTGGCTGACCCCCAGCCGAGCGGCTCCAGTACCT 1020
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1685  GGTGGGTATGCTTCGGCAGAGCCCGCCGCCCGACCCACCCTGGA  1744

Db  1021  GGTGGGTATGCTTCGGCAGAGCCCGCCGCCCGACCCACCCTGGATCCTGACCGGCTGCCAACT 1080
           |||||||||||||||||||||||||||||||||||
Qy  1745  GCCCGCCCCCATGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGGGCCTGCA 1804

Db  1081  GCCCGCCCCCATGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGGGCCTGCA 1140
           |||||||||||||||||||||||||||||||||||||||||
Qy  1805  GACCCTCAGTCTGGCCTCTGTGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGT 1864

Db  1141  GACCCTCAGTCTGGCCTCTGTGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGT 1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1865  GAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGGCCACCCACAACC 1924

Db  1201  GAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGGCCACCCACAACC 1260
           |||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1925  TCCCAAGGAACTCATCTCGACCTTCTGAGGCCCTGTGGCCAGAGAGGGTGGAAGACCC 1984

Db  1261  TCCCAAGGAACTCATCTCGACCTTCTGAGGCCCTGTGGCCAGAGAGGGTGGAAGACCC 1320
           |||||||||||||||||||||||||||||||||||||||||||||
Qy  1985  TAGTCAAAGTCCCTGTGGAGAGAACGGCCCCATTCCAAGGGCAGGAGGATATTGCTCCGG 2044

Db  1321  TAGTCAAAGTCCCTGTGGAGAGAACGGCCCCATTCCAAGGGCAGGAGGATATTGCTCTCG 1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2045  CCTTTGGAAACTTTGAGCCGAGAGGCGCGAGACAGGCATGTGGAGGCCCAGACACGG 2104

Db  1381  CCTTTGAAAACTTTGAGCCCGAGAGGCGCGAGACAGCAGCATGTGGAGGCCCAGACACGG 1440
           |||||||||||||||||||||||||||||||||||||||||||||||
Qy  2105  CACCCTGCCCGTCCAGGACAGCGCCAGGACTGCCCCCTCTCTCCACACCTGGGTACCC 2164

Db  1441  CACCCTGCCCGTCCAGGCACAGCGCCAGGACCTGCCCTCTCTCCACACCTGGGGTACCC 1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2165  CTTCTCCCCAGCCCCACCACTACTCCACCCACCTTCCTCCTCCTGAGACCCTCCAGCCAT 2224

Db  1501  CTTCTCCCCAGCCCCACCACTACTCCACCCACCTTCCTCCTCCTGAGACCCTCCAGCCAT 1560
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2225  TCCCCTTGAAAACACCCCCGACCCCAGCCACAATAATGACAGCGGAGCGAGCTCCAATAA 2284

Db  1561  TCCCCTTGAAAACACCCCCGACCCCAGCCACAATAATGACAGCGGAGCGAGCTCCAATAA 1620
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2285  CTAAGCACCTACCCGGCGGCAGAATAACCCTTCACTGCCTAATCCCATCTGCAGTGTGG 2344

Db  1621  CTAAGCACCTACCCGGGGGCAGAATAACCCTTCACTGCCTGATCCCCATCTGCAGTGTGG 1680
           ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2345  CCCACAGCCCCCAGAACCATGCCCACATAGACTGGAGGTAGGCAGTCACCGTCCCTCC 2404

Db  1681  CCCACAGCCCCCCAGAACCCCCACATAGACTGGAGGTAGGCAGTTCACCGTCCCTCC 1740
           ||||||||||||||||||||||||||||||||||||||||||||
Qy  2405  CTGTTAGGAATGAGACCATCCCTGAGGCTATGGCCCACAGGCGTCCAGTGTCTG 2464

Db  1741  CTGTTAGGAATGAGACCATCCCTGAGGCTATGGCCCACAGGCGTCCAGTGTCTG 1800
           ||||||||||||||||||||||||||||||||||||||||||||
Qy  2465  AGATCTTTGGGAGGGACTAGGGCAGGTGGAGCAGCGCAGGAGAACCCCGTGCTGGGTG 2524

Db  1801  AGATCTTTGGGAGGGACTAGGGCAGGTGGAGCAGCGCAGGAGAACCCCGTGCTGGGTG 1860
           ||||||||||||||||||||||||||||||||||||||||||||||
Qy  2525  GGAAGCATGACCACACATGGTGGGTGAGCAGCCCCCATGCACTGACGGTAAATTCCCCGTG 2584

Db  1861  GGAAGCATGACCACACATGGTGGGTGAGCAGCCCCCATGCACTGACTGATGGTAAATTCCCCGTG 1920
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2585  GACTCATTTCTGTTGGTTTCTATTACACCTGGCCAGCCGTGTACATACAGGTCGGTGC 2644

Db  1921  GACTCATTTCTGTTGGTTTCTATTACACCTGGCCAGCCGTGGTACAATACAGTTCGGTGC 1980
           ||||||||||||||||||||||||||||||||||||||||||||||||
Qy  2645  TCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2689

Db  1981  TCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 2025
```

RESULT 5

```
Qy  187 GCGGCCGGGTGCTAGGCGGCCTGCTGAGTAAGGCGGCCTGCTGCCCAGGCCCTCCTGCTGG 246
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  911 GCGGCCGGGTGCTAGGCGGCCTGCTGAGCGAGGCGGCCTGCTGCCCACGGCCCTCCTGCTGG 970

Qy  247 TGACCACGGCCGCGCCGCCGCCGCCGCCCGGGAGGCTGCAGGCGCCGCCCTGTTCCCGAGTCGG 306
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  971 TGACCACGGCCGCGCCGCCGCCGCCCGGGAGGCTGCAGGCGCCGCCCTGTGTTCCCGAGTCGG 1030

Qy  307 CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAGAAGTATTCTACAAGTTCTTCCCGGG 366
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1031 CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAGAAGTATTCTACAAGTTCTTCCCGGG 1090

Qy  367 ATGAGGAGGAGGCCGGAGCGCCCTACCGCGTTCGTGAAGGAGAACGAGCGCTGTCGCGGC 426
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1091 ATGAGGAGGAGGCCGGAGCGCCCTACCGCGTTCGTGAAGGAGAACGAGCGCTGTTCGCGCC 1150

Qy  427 TGTGCTTGCTGCCCTTCGTGTGTGGATCGTGTGCACCGCTGCGCCAGCAGCTGGGAGC 486
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1151 TGTGCTTGCTGCCCTTCGTGTGGCTGGATCGTGTGCACCGCTGCGCCAGCAGCTGGGAGC 1210

Qy  487 TCGGTCGGGACCTGCGCGCAGTCCAAGACCACACACGTCAGTGTACCTGCTTTTCATCA 546
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1211 TCGGTTCGGGACCTGCGCGCAGTCCAAGACCACACACGTCAGTGTACCTGCTTTTCATCA 1270

Qy  547 CCAGGCGTTCTGAGCTCGGCTTCGGTAGCCGAGGCGCCCGGTTGCAGGGCGACCTGGCGA 606
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1271 CCAGGCGTTCTGAGCTCGGCTTCGGTAGCCGAGGCGCCCGGTTGCAGGGCGACCTGGCGA 1330

Qy  607 ATCCTGTGCCGCCCTGGCCCGGAGGCGTCCTCGGACGCAGGGGCAGTTGCCGGAGAGGS 666
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1331 ATCCTGTGCCGCCCTGGCCCGGAGGCGTCCTCGGACGCAGGGGCAGTTGCCGGAGAGG 1390

Qy  667 AACTGGAGCAACTGAGCTTGTGGCTCCAAAGTGCAGCGCCGTTTCTCAGCAGCAAAAGG 726
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1391 AACTGGAGCAACTGAGCTTGTGGCTCCAAAGTGCAGCGCCGTTTCTCAGCAGCAAAAGG 1450

Qy  727 AGCTGCGCGGGGTGCTGAGACAGAGGTCACCTACCAGTTCATCGACCAGCAGCTTCCAGG 786
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1451 AGCTGCGCGGGGTGCTGAGACAGAGGTCACCTACCAGTTCATCGACCAGCAGCTTCCAGG 1510

Qy  787 AGTTCCTCGCGGACTGTCCCACTGCCTGGAGGACGCGGCCCAGGACCGCGGCGTG 846
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1511 AGTTCCTCGCGGACTGTCCCACTGCCTGGAGGACGCGGCCCAGGACCGCGGCGTG 1570

Qy  847 GCGGCGGTTGGGACACTCCTGCGTGGGGACACGCCAGCCGCACAGCCCACTTGGTGCTCACCA 906
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1571 GCGGCGGTTGGGACACTCCTGCGTGGGGACACGCCAGCCGCACAGCCCACTTGGTGCTCACCA 1630

Qy  907 CGCGGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGCGGACTGCGGCGGAGCGGCCCACTTCG 966
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1631 CGCGGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGCGGACTGCGGCGGAGCGGCCCACTTCG 1690

Qy  967 GCTGCATGGTTTCAGAGCGTGTGAAGCAGCAGGGGCCCTGCCGGTGGGTGCAGCAGGGAC 1026
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1691 GCTGCATGGTTTCAGAGCGTGTGAAGCAGCAGGGGCCCTGCCGGTGGGTGCAGCAGGGAC 1750

Qy 1027 AGGGCTGCCCCGAGGGCACCCAGAGGTGGCACCAGAGGTGACCCAGAGGGCTCGAGGACACCG 1086
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1751 AGGGCTGCCCCGGAGTGGCACCAGAGGTGGCACCAGAGGTGACCCAGAGGGCTCGAGGACACCG 1810

Qy 1087 AAGAGCCAGGAGGAGGAGGAGGAGGAGGCCCAACTACCCACTGGAGTTGCTGTACT 1146
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1811 AAGAGCCAGGAGGAGGAGGAGGAGGAGGCCCAACTACCCACTACCCACTGGAGT TGCTGTACT 1870

Qy 1147 GCCTGTACGAGACGACCGACGCGCGCTTTGCGCCAAGCCCTGTGCCCGGTTCCCGGAGC 1206
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1871 GCCTGTACGAGACGACGCGCGCTTTGTGCGCCAAGCCCTGTGCCCGGTTCCCGGAGC 1930

Qy 1207 TGGCGCGCCGAGGTGCGCGCGCTTCTCCGCGATGCAGGGCGACCTGGCGCTTGCTGAGCTGCGTGA 1266
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1931 TGGCGCTGCAGCGGAGTGCGCGCTTCTCCCGCGCATGGACGTGGCGTTCTGAGCTGCGTGA 1990

Qy 1267 GGTGCTGCCCTGCTGGACAGCACTGCGGCGCTGATCAGCTGCGAGATTGGTTGCTGCGCAGG 1326
```

```
Db 1991 GGTGCTGCCCTGCTGGACAGGCACTGCGGCGCTGATCAGCTGCGAGATTGGTTGCTGCGGAGG 2050

Qy 1327 AGAAGAAGAAGAAGAGCTGGGGAAGCGGCTCCAGGCCAGCTGATCAGCTGCAGATTGGTTGCTGCGGG-- 1383
            ||||||||||||||||||||||||||||||||||||||||||||||||||||| : :||
Db 2051 AGAAGAAGAAGAAGAGCTGGGGAAGCGGCTCCAGGCCAGCTGGGCGGCAGCAGTT 2110

Qy 1384 CTCAAGGCACCACAAAACAACTGCCAGCTCCCTTCTTCATCCACTCTTTCAGGCAATGA 1443
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2111 CTCAAGGCACCACAAAACAACTGCCAGCTCCCTTCTTCATCCACTCTTTCAGGCAATGA 2170

Qy 1444 CTGACCCACTGTGCCATCTGAGCAGCCTCACGTGCTCCACTGCAAACTCCCTGACGCGG 1503
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2171 CTGACCCACTGTGCCATCTGAGCAGCCTCACGTGCTCCACTGCAAACTCCCTGACGCGG 2230

Qy 1504 TCTGCCGAGACCTTCTGAGGGCCTGAGGGCCAGCGCCCCCGCACTGACGGAGCTGGGCCTCC 1563
            ||||||||||||||||||||||||||||||||||||||||||||| |||||||||
Db 2231 TCTGCCGAGACCTTTCTGAGGGCCTGAGGGCCAGCGCCCCCGCACTGACGGAGC TGGGCCCC 2290

Qy 1564 TCCACACAGGTCAGTGAGCGGGACTGGTTGAGTGAGTGCTGTATGGTGAGTGAGCGTCCGGGCCTGC 1623
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2291 TCCACACAGGTCAGTGAGCGGGACTGGTTGAGTGAGTGCTGTATGGTGAGTGAGC GTCCGGCCGC 2350

Qy 1624 AGTGCAGGTGCAGACCGTCAGGTAGCAGTTGCTGCTACAGCGAGGGGCTCCAGTACC 1683
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2351 AGTGCAGGTGCAGACCGTCAGGTAGCAGTTGCTGCTACAGCGAGGGGCTCCAGTACC 2410

Qy 1684 TGGTGGGTATGCTTCGGCGAGAGCCCCGGCCTCCTCGATCTCAGCGGCTGCCAAC 1743
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2411 TGGTGGGTATGCTTCGGCGAGAGCCCCGGCCTCCTCGATCTCAGCGGCTGCCAAC 2470

Qy 1744 TGCCCGCCCCATGGTGACCTACCTGTGTGCAGTCTCTGCAGCACCAGGATCGGGCCTGC 1803
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2471 TGCCGCCCCCATGGTGACCTACCTGTGTGCAGTCTCTGCAGCACCAGGATGCGGGCCTGC 2530

Qy 1804 AGACCCTCAGTCTGGCCTTCTGTGGAGCGGAGCAGTCACTACGAGAGAGCTTCAGGCTG 1863
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2531 AGACCCTCAGTCTGGCCTTCTGTGGAGCGGAGCAGTCACTACAGGAGAGCTTCAGGCTG 2590

Qy 1864 TGAAGAGCAAAGCCGATCTCGTCATCACACACCGAGGCGCCACCACAC 1923
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2591 TGAAGAGCAAAGCCGATCTCGTCATCACACACCGAGGCGCCACCACAC 2650

Qy 1924 CTCCCAAGGAACTCATTCGACCTTCGA 1952
            |||||||||||||||||||||||||||
Db 2651 CTCCCAAGGAACTCATTCGACCTTCTGA 2679
```

RESULT 6
AY154461

LOCUS       AY154461                2679 bp    mRNA     linear   PRI 20-FEB-2003
DEFINITION  Homo sapiens NALP6 (NALP6) mRNA, complete cds.
ACCESSION   AY154461
VERSION     AY154461.1  GI:28436365
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2679)
  AUTHORS   Tschopp,J., Martinon,F. and Burns,K.
  TITLE     NALP6: a novel protein family involved in inflammation
  JOURNAL   Nat. Rev. Mol. Cell Biol. 4 (2), 95-104 (2003)
   MEDLINE  22451042
    PUBMED  12563287
REFERENCE   2  (bases 1 to 2679)
  AUTHORS   Martinon,F., Hofmann,K. and Tschopp,J.
  TITLE     Direct Submission
  JOURNAL   Submitted (25-SEP-2002) Institute of Biochemistry, University of
            Lausanne, ch. des Boveresses 155, Epalinges, VD 1066, Switzerland
FEATURES             Location/Qualifiers
     source          1..2679
                     /organism="Homo sapiens"

```
gene      1..2679
          /db_xref="taxon:9606"
          /gene="NALP6"
CDS       1..2679
          /gene="NALP6"
          /note="member of the NALP protein family involved in
          inflammation; contains NACHT, LRR and PYD domains"
          /codon_start=1
          /product="NALP6"
          /protein_id="AAO18157.1"
          /db_xref="GI:28436366"
          /translation="MDQPEAPCSSTGPRLAVARELLAALELSQELKRFRKLRDV
          GPDRGRSIPWGRLERADAVDLAEQLAQFYGPEPALEVARKTLKRADARDVAAQLQERRL
          QRLGLGSSTLLSVSEYYKKYREHVLQLHARVKERNARSVKITKRFTKLIAPESAAPE
          EALGPAEPEPGRARRSDTHTFNRLFRRDEEGRRPLTVVLGGKTMAAAKKILYD
          WAAGKLIYGQVDFAFFMPCGELLERPGTRSLADLILDQCPDRGAPVPQMLAQPQRLLF
          ILDGADELPALGPEAAPCTDFFEAASGARVLGGLLSKALLPTALLLVTTRAAAPGRL
          QGRLCSPQCAEVRGFSDKDKKKYFYKFRDERRAEBRAYRFVKENETLFALCFVPFVCW
          IVCTVLRQQLELGRDLSRTSKTTTSVVLLFITSVLSSAPVADGPRLQGDLRNLCRLAR
          EGVLGRRAQFAEKELEQELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFLAA
          LSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM
          VSERVKQEALRNVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELLYC
          LYETQEDAFVRQALCRFPELALQQRVFRCRMDVAVLSYCVRCCPAGQALRLISCRLVAA
          QEKKKKSLGGSSSGTTKLQLPASLLHPLFQAMTDPLCHLSSLTLSHCKL
          PDAVCRDLSEALRLAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDFQ
          RGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQ
          SLQELQAVKRAKPDLVITHPALDGHPQPPKELISTF"
BASE COUNT      458 a      878 c      934 g      409 t
ORIGIN

Query Match        71.6%;  Score 1925;  DB 9;  Length 2679;
Best Local Similarity  99.6%;  Pred. No. 2.5e-291;
Matches 1941;  Mismatches 0;  Conservative 0;  Indels 5;  Gaps 3;
```

```
Qy     7  CGTGGGCGGCAGCCTGGCTGACCTGATCCTGACCCAGTGCCCCGACCGGCGGCGCGCCGG   66
          |||  ||||| |||||||||||||||||||||||||||||||||||||||||||||||||
Db   731  CGGGCACGCGCAGCCTGGCTGACCTGATCCTGACCCAGTGCCCCGACCGGCGGCGCGCCGG  790

Qy    67  TGCCCGAGATGCTGGCCCAGCCGCAGCGCCTGCTCTTCATCCTGGACGGCGCGGACGAGC  126
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   791  TGCCCGAGATGCTGGCCCAGCCGCAGCGCCTGCTCTTCATCCTGGACGGCGCGGACGAGC  850

Qy   127  TGCCGGCGCTGGGGGCGCCCCGAGGCCGGGCCGGCCTCCGAGGCGGCGGACGG        186
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   851  TGCCGGCGCTGGGGGCGCCCCGAGGCCGGGCCGGCCTCCGAGGCGGCGGACGG        910

Qy   187  GGGCGCGGGTGCTAGGCGGGCTGCTGAGTAAGGGCCTGAGCCCCTCCGGG        246
          ||||||||||| | |||||||||||||||||||||||| | ||||||||
Db   911  GGGCGCGGGTGCTAGGCGGGCTGCTGAGCAAGGGCCTGAGCCCCTCGTGG        970

Qy   247  TGACCACGGCGGCCGGCCCCGGCCCCGGGAGGCGCTGGGGAGAGCGCTGGG        306
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db   971  TGACCACGGCGGCCGGCCCCGGCCCCGGGAGGCGCTGGGGAGAGCGCTGGG       1030

Qy   307  CCGAGGCGCGGCCGCCGGGTTCTCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG  366
          ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1031  CCGAGGCGCGGCCGCCGGGTTCTCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 1090

Qy   367  ATGAGAGGAGGGCCGAGGCGCGCCACCGCTGGGATGTGTGCTGGAC        426
          |||||||||||||||||||||||||||||||||||||||||||||
Db  1091  ATGAGAGGAGGGCCGAGGCGCGCCACCGCTGGGATGTGTGCTGGAC       1150

Qy   427  TGTGCTTCGTGCCCTTCGTGTGTCTGGATGTGTGCACCGTGCTGCGCCAGAGCTGGAGC  486
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1151  TGTGCTTCGTGCCCTTCGTGTGTCTGGATGTGTGCACCGTGCTGCGCCAGAGCTGGAGC 1210

Qy   487  TCGGTTCGGGACCTGTCCGAAGCGTCCGAAGACCACCACCAGTCAGTGTACCTGCTTTCATCA  546
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1211  TCGGTTCGGGACCTGTCCGAAGCGTCCGAAGACCACCACCAGTCAGTGTACCTGCTTTTCATCA 1270

Qy   547  CCAGCGTTCTGAGCTCGGCTCCGTAGCGCACGGGCCCCGGTTCGAGCGGCGACCTGCCA
          ||||||||||||||||||||||||||||||||
```

```
         CCAGCGGTTCTGAGCTCTGGCTTCCGGCTAGCGGCCCTCCGGTAGCGGCAGGGGCGCAACTTGGCCGAGAGG 1330
1271     CCAGCGGTTCTGAGCTCTGGCT...                                               666
Qy       ATCTGTGCCGCCTGCCTGCGCCTGCTCGGAGTCAGCACTCGGCTGGTGCTGCTGCAGGGGCAGTTGCCGAGAAGG
         ||||| ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1331   ATCTGTGCCGCCTGCCTGCGCCTGGCTCTCGGAGTCAGCACTCGGTCTGCTGCTGCAGGGGCAGTTGCCGAGAAGG 1390
Qy  726   AACTGGAGCAACTGGAGCTTCGTGCTGTGTCGCAGACGCGCGTTTCTCAAAGTCAGACGCGTCCGAAAAGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 667    AACTGGAGCAACTGGAGCTTCGTGCTGTGTCGCAGACGCGCGTTTCTCAAAGTCAGACGCGTCCGAAAAGG 1450
Qy 786    AGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 727    AGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 1510
Qy 846    AGTTCCTGCGCGACTGCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACGCGCGGCGTG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 787    AGTTCCTGCGCGACTGCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACGCGCGGCGTG 1570
Qy 906    GCGGCGGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTGGTGCTCACCA
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 847    GCGGCGGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTGGTGCTCACCA 1630
Qy 966    CGCGCTTCCTCTTCGGACTGCTGAGCGCCGAGCGGATGCGCGACATCGAGCGCCACTTCG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 907    CGCGCTTCCTCTTCGGACTGCTGAGCGCCGAGCGGATGCGCGACATCGAGCGCCACTTCG 1690
Qy 1026   GCTGCATGGTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGTGGGTGGGGTGCAGGGAC
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 967    GCTGCATGGTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGTGGGTGGGGTGCAGGGAC 1750
Qy 1086   AGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGGGGGCCAAGGGCTCGAGGACACCG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1027   AGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGGGGGCCAAGGGCTCGAGGACACCG 1810
Qy 1146   AAGAGCCAGGAGGAGGAGAGGCTGGGAGTACACCCACTGGAGTTGCTGTACT
          ||||||||||||||||||||||||||||||||||||||||||||||
Db 1087   AAGAGCCAGGAGGAGGAGAGGCTGGGAGTACACCCACTGGAGTTGTGTAC 1870
Qy 1206   GCCTGTACGAGACGCAGGAGGACACCGTTGTGCGCCAGGCCTTCCCGGAGC
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db 1147   GCCTGTACGAGACGCAGGAGGACACCGTTGTGCGCCAGGCCTTCCCGGAGC 1930
Qy 1266   TGGCGCTGCAGCGAGTGGCGCTTCTGCCGCCATGGACGTGGCTGT CTGAGCTACTGCGTGA
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1931   TGGCGCTGCAGCGAGTGGCGCTTCTGCCGCCATGGACGTGGCTGT CTGAGCTACTGCGTGA 1990
Qy 1326   GGTGTGCCCTGCTGGACAGGCACTGCGGCTGACTGCGGCTCAGTTGGTTGCTGCGCGCAGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1991   GGTGCTGCCCTGCTGGACAGGCACTGCGGCTGACTGCGGCTCAGTTGGTTGCTGCGCGCAGG 2050
Qy 1383   AGAAGAGAAGAGCCTGGGAGGCGCTGGGGAAGCGGCTCAGGCGGCTGGGG--GCAGTT
          ||||||||||||||||||||||||||||| ||||||||||||||| |||
Db 2051   AGAAGAGAAGAGCCTGGGAGGCGCTGGGGAAGCGGCTCAGGCGGCTGGGGCGGCAGTT 2110
Qy 1443   CTCAAGGCACCACAAAAACACTGCCAGCCTCCTTCTTCATCCACTCTTCGAAACTTCCCGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2111   CTCAAGGCACCACAAAAACACTGCCAGCCTCCTTCTTCATCCACTCTTCGAAACTTCCCGG 2270
Qy 1503   CTGACCCACTGTGCCATCTGACCGCCTCACGCTGTCCCACTGCAAACTCCCGACGCGG
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2231   CTGACCCACTGTGCCATCTGACCGCCTCACGCTGTCCCACTGCAAACTCCCGACGCGG 2230
Qy 1563   TCTGCGACCTTTCTGAGGTGTGGATCGTGTGCTGCCCTGCGCGCAGCGCTGGAGC
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2231   TCTGCGACCTTTCTGAGGTGTGGATCGTGTGCTGCCCTGCGCGCAGCGCTGGAGC 2290
Qy 1623   TCCAACACAGGCTCAGTGAGCGGACTGCGTATGCGCGTATGCGCCTAGCCTGGCCGC
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2291   TCCAACACAGGCTCAGTGAGCGGACTGCGTATGCGCGTATGCGCCTAGCCTGGCCGC 2350
Qy 1683   AGTGAGGGTGCAGACGTGCAGGTTCAGGGTACACCCCAGTAGGGCTGCCCAGTACC
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 2351   AGTGAGGGTGCAGACGTGCAGGTTCAGGGTACACCCCAGTAGGGCTGCCCAGTACC 2410
```

```
Qy   1031 CCGAGGTGCGCGGTTCTCCGACAGGACAAGAGGAAGAAGTATTTCTACAGTTCTTCCGGG 1090
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    :743 TGGTGGGTATGCTTCGGCAGAGCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAAC :743

Qy    426 ATGAGAGGAGGGCCGAGCCGCCTACCGTTCGTGAAGGAGAACGAGACGCGTGTTCGCGC  426
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2411 TGGTGGGTATGCTTCGGCAGAGCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAAC 2470

Qy   1091 ATGAGAGGAGGGCCGAGCCGCCTACCGTTCGTGAAGGAGAACGAGACGCGTGTTCGCGC 1150
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1744 TGCCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGCCTGC 1803

Qy    486 TGTGTTCGTGTGCCCTTCGTGCTGTGGATCGTGTGCCAGCGTGCGCGCCAGCAGGTGGAGC  486
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2471 TGCCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGCCTGC 2530

Qy   1151 TGTGTTCGTGTGCCCTTCGTGCTGTGGATCGTGTGCCAGCGTGCGCGCCAGCAGGTGGAGC 1210
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1804 AGACCCTCAGTCTGGAGCTGGGACCTGTCGGCAGTCGGAGCAGTCACTACAGAGGCTTCAGGCTG 1863

Qy    546 TCGGTCGGGACCTGTCGGCAGTCGGAGCAGTCACTACAGAGGCTTCAGGCTGTACCTGCTTTCATCA  546
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2531 AGACCCTCAGTCTGGCCTCTGTGGAGCTGGGACCTGTCGGCAGTCGGAGCAGTCACTACAGAGGCTTCAGGCTG 2590

Qy   1211 TCGGTCGGGACCTGTCGGCCTCTGTGGAGCTGGGACCTGTCGGCAGTCGGAGCAGTCACTACAGAGGCTTCAGGCTG 1270
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1864 TGAAGAGAGCAAAGCCGGATCCGGATCCGGTCATCACACACACCACCACCACCCACCAAC 1923

Qy    606 CCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGCGTCCTGCGACGAGGGCGCAGGGCCAGTTGCCGAGAAGG  606
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2591 TGAAGAGAGCAAAGCCGGATCCGGATCCGGTCATCACACACACCACCACCACCCACCAAAC 2650

Qy   1271 CCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGCGTCCTGCGACGAGGGCGCAGGGCCAGTTGCCGAGAAGG 1330

Qy    666 ATCTGTGCGCGCCTGGCCCGCGAGGGCGTCCTCGGAGCAGGGCCGACTGGCCAGGGCA  666
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1924 CTCCCAAGGAACTCATCTCGACCTTCTGA 1952

Qy   1331 ATCTGTGCGCGCCTGGCCCGCGAGGGCGTCCTCGGACGCGGCAGGGCCGCAGGTTGCCGAGAAGG 1390
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2651 CTCCCAAGGAACTCATCTCGACCTTCTGA 2679

Qy    726 AACTGGAGCAACTGGAGTTCGTGGCTCCAAACTGCAGACGCTGTTTCTCAGCAGAAAAAAGG  726

Qy   1391 AACTGGAGCAACTGGAGCTTCGTGGCTCCAAACTGCAGACGTGCAGACGCTGTTTCTCAGCAGCAAAAAAGG 1450

Qy    786 AGCTGCCGGGCGGTGCTGCTGAGACAGAGTTCACCAGTTCATCGACCAGAGCTTCCAGG  786

Qy    727 AGCTGCCGGGCGGTGCTGCTGAGACAGAGTTCACCAGTTCATCGACCAGAGCTTCCAGG  727

Qy   1451 AGCTGCCGGGCGGTGCTGCTGAGACAGAGTTCACCAGTTCATCGACCAGAGCTTCCAGG 1510

Qy    846 AGTTCCTCGGCACGTCCTCACCTGCTGGAGGACGAGCGGGGGTGCCCAGGACCCGCGGCTG  846

Qy   1511 CGGGCCGTTGGGACACTCCTCGCGTGGGGACGCCCAGCAGGGCGGGTGCCCAGGACCGCGGCTG 1570

Qy    906 GCGGCGTTGGGACACTCCTCGGTGGGGGACGCGGGACACTGGTCGTGGGTGCTCACCA  906

Qy   1571 GCGGGCGTTGGGACACTCCTCGCGTGGGGACGCGGGACACTGGTCGTCACCA :630

Qy    966 CGGGCCTTCCTCTTCGGACTGCTCGGAGACTGCTCGAGCGGCCGATGCGGGACATCGAGCGGCCGCGGCTG  966

Qy   1631 CGGGCCTTCCTCTTCGGACTGCTCGGAGACTGCTCGAGCGGCCGATGCGGGACATCGAGCGGCCGCGGCTG :690

Qy   1027 GCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGGCCTGCGGGTGGTGTCAGGGACGCAGGGAC 1086
Db   1750 GCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGGCCTGCGGGTGGTGTCAGGGACGCAGGGAC 1750

Qy   1086 AGGGCTGCCCCGAGTGCACCAGGACCTCGAGGGTGCGAGGAGGGCCAAAGGGTCGAGGACACCG 1086

Db   :751 AGGGCTGCCCCGAGTGGTGCACCCAGAGGTGCACCCAGAGGTGACCGAGGGTGCGAGGACACCG :810

Qy   1087 AAGAGCCAGAGGAGGAGGAGGAGAGAGGAGGAGGAGCCCAACTACCACTGGAGTTGCTGTACT 1146

Db   1087 AAGAGCCAGGAGGAGGAGGAGGAGGAGAGAGGAGGAGGAGCCCAACTACCACTGGAGTTGCTGTACT  66

Qy   1811 AAGAGCCAGGAGGAGGAGGAGGAGGAGGAGGAGCCCAACTACCACTGGAGTTGCTGTACT 1870

Qy   1147 GCCCTGTACGAGAGACGCAGGGGACGCGTTTGTGCGCCAAGCCCGTTCCCGGTTCCCGGGGAGC 1206

Db   :871 GCCTGTACGAGAGACGCAGGGGACGCGTTTGTGCGCCAAGCCCGTTCCCGGTTCCCGGAGC 1930

Qy   1207 TGGCGCTGCAGCGAGTGCGGCTTCTGCCGGCCACTGGGCACTGGCGTGTTCTGAGGCGGTGGTGA 1266

Db   1931 TGGGCGCTGCAGCGAGTGCGGCTTCTGCCGCCGCATGGACGTGCGTGTTCTGAGACGGTGGTGA 1990

Qy   1267 GGTGCTGCCCTGCTGGGGTGCTGAGTAGGCCCTGCTGAGTCAGGGCACTGCGCAGATTCGGTTGTGCGCCAGG 1326

Db   1991 GGTGCTGCCCTGCTGGGTGCTGAGTAGGCCCTGCTGAGTCAGATTCACTGCGCAGATTGGTTGTTGCGCGAGG 2050

Qy   1327 AGAAGAAGAGAGACGAGCCTGGGGGAAGCGGCTCAGGGGAAGCGGCTCCCAGCCAGTCTCC 1386

Db   2051 AGAAGAAGAGAGAAGAGAGCGACCTTGGGGGAAGCGGCTCCCAGCCCAGCCAGCTGC 2110

Qy   1387 AAGGCACCACAAAACAACTGCCACCCTTCATCCACCTCTT---------------- 1433

Db   :111 TGGGGACCCCAACTGGCTCCAGGAGTACCCTTTGACCACACCCTGTGTGACATCGCCCCA 2170
```

```
RESULT 7
AX088215
LOCUS       AX088215                5553 bp    DNA     linear       PAT 17-MAR-2001
DEFINITION  Sequence 10 from Patent WO0114564.
ACCESSION   AX088215
VERSION     AX088215.1    GI:13397126
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Peyman,J.A., Green,C.D., Hsu,A., Browning,J.A. and Carulli,J.
  TITLE     Polynucleotides expressed in activated t-lymphocytes and proteins
            encoded thereby
  JOURNAL   Patent: WO 0114564-A 10 01-MAR-2001;
            Curagen Corporation (US); BIOGEN, INC. (US);
FEATURES             Location/Qualifiers
     source          1..5553
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      912 a    1856 c    1879 g    906 t
ORIGIN
```

```
Query Match          58.4%;   Score 1571.2;  DB 6;  Length 5553;
Best Local Similarity  89.6%;   Pred. No. 4.6e-236;
Matches 1772;  Conservative   0;  Mismatches  38;  Indels  168;  Gaps   1;
```

```
Qy      7 CGTGGGGCGCGGCCAGCCTGGCT3ACCTGGCT3ACCTGGATCCGATCCTGGACACCAGTGCCCCAGGCACCCCGACCAGTGCCCCCAGACGCCCGG  66
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731 CGGGCACGGCGGCCAGCCTGGCT3ACCTGGCT3ACCTGGATCCGATCCTGGACACCAGTGCCCCCAGCACCCCCGACAGGCGCGGGCGCGGCCCG  790

Qy     67 TGCGGCGCAGATGCTGGCCGCGCCCAGGGGGCGCGCCGAGCGGCGCCGAGCCTCGAGGCGGCGGGAGAGG 126
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791 TGCCCGACAGATGCTGGCCGCGCCCAGCGCCGCGGCGCCGAGCGGCGCGGCGGGAGAGG  850

Qy    127 TGCCGGCGCGTGGGGGGGGGCCTCGGGGGGGCCCTCGAGGCGCGGCGAGGG 186
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851 TGCCGGCGCGTGGGGGGGGGCCTCGGGGGCCCCTTCGAGGCGGCGGCGAGGG 910

Qy    187 GCGGCGCGGGTGCTGAGGCGGGTGCTGAGTAAGGCCCTGCTGTCCCACGCGCCTCTGCTG 246
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911 GCGCGGCGGGGTGCTGAGGCGGTGCTGAGGCAAGCCCTGCTGCTGCTGTCCCCACGCGCCTCCTGCTG  970

Qy    247 TGACCACGCGCGCCGCCGCCCGCGCCCCCCGGGAGGCTGCAGGGCGGCTGTTCCCCCAGTGGCG 306
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971 TGACCACGCGCGCGCGCCCCCCCGCGCCCCCCGGGAGGCTGCGCAGGCGCCCGTGTTCCCCGCAGTGGCG 1030

Qy    307 CCGAGGTGCGCGCGGCTTCTCCGACAGGACAAGAGGAAGAAGTATTTCTACAGTTCTTCCGGG  366
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   :111 TGGGGACCCCAACTGGCTCCAGGAGTACCCTTTGACCACACCCTGTGTGACATCGCCCCCA 2170
```

```
Qy   1434  -----------------------------------------------------------  1434

Db   2171  CACCTCCCACAGACCCTCGGCTCCTCCAGGCAGGCTTTTGCCAGGAGTTCCTTTGAATA   2230
                                                                         
Db   1434  -----------------------------------------------------------  1434

Qy   2231  TAGTCCAATTCAGCCCCTGCCCAGGGGGCTTGGCATCTGTTGAGAGGATGAATGTCACGG   2290
                                                                         
Db   1434  -----------------------------------------------------------  1434

Qy         TGTTGGCAGGGGCTGGGCCTGGGACCCTGGGACCCAAAGACCCATGCAATGACTGACCCACTGTCGCC
Db   2291  ||||||||||||||||||||                                                
                      CAGGCAATGACTGACCACCACTGTGCC                           2350

Qy   1459  ATCTGAGCAGCCCTCACGCTGTCCCACTGCAAACTCCTGACGCGGTCTGCCGAGACCTTT  1518
           |||||||||||||||||||||||||||| |||||||||||||||||||||||||||||||
Db   2351  ATCTGAGCAGCCCTCACGCTG-CCCACTGCAAACTCCTGACGCGGTCTGCCGAGACCTT-  2410

Qy   1519  CTGAGGGCCCTGAGGGCAGCCCCACTGACGGAGCTGGCCCTCGCCCTCCTGCCAACAGGGTCA  1576
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2411  CTGAGGGCCCTGAGGGCAGCCCCACTGACGGAGCTGGCCCTCGCCCTCCTGCCAACAGGGTCA  2470

Qy   1579  GTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCCTAGCCTGGCCGCCCAGTGCAGGGTGCAGA  1638
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2471  GTGAGGCAGGACTGCGTATGCTGAGTGAGGGCCCTAGCCTGGCCGCCCAGTGCAGGGTGCAGA  2530

Qy   1639  CGGTCAGGGTACAGCTGCCTGACCCCCAGCGCGAGGGCTCCAGTACCTGGTGGGTATGCTTC  1698
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2531  CGGTCAGGGTACAGCTGCCTGACCCCCAGCGCGAGGGCTCCAGTACCTGGTGGGTATGCTT-C  2590

Qy   1699  GGCAGAGCCCCGCCCCTGACCACCCCTGACCTGCACGCGGATCTCAGGCGGGCCCTGCTGGG  1758
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   2591  GGCAGAGCCCCGCCCCTGACCACCCCTGACCTGCACGCGGATCTCAGGCGGGCCCTGCTGGG  2650

Qy   1759  TGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGGCCCTGCGAGACCCTCAGTCT  1816
           |||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db   2651  TGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGGCCCTGCGAGACCCTCAGCCT  2708
```

RESULT 8
AX459875

LOCUS       AX459875               5556 bp    DNA     linear   PAT 08-JUL-2002
DEFINITION  Sequence 18 from Patent WO0240668.
ACCESSION   AX459875
VERSION     AX459875.1  GI:21725646
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tschopp,J. and Martinon,F.
  TITLE     Proteins and dna sequences underlying these proteins used for
            treating inflammations
  JOURNAL   Patent: WO 0240668-A 18 23-MAY-2002;
            Apotech Research and Development Ltd. (CH)
FEATURES             Location/Qualifiers
     source          1..5556
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="NALP6/PY9.dna"
BASE COUNT      914 a   1856 c   1879 g    907 t
ORIGIN

Query Match  58.4%;  Score 1571.2;  DB 6;  Length 5556;
Best Local Similarity  89.6%;  Pred. No. 4.6e-236;
Matches 1772;  Conservative  0;  Mismatches  38;  Indels  168;  Gaps  1;

```
Qy      7  CGTGGGCGCGCAGCCTGCCCAGCCTGACCTGACCCGCCCCGACCCGCGGCGG   66
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    731  CGGGGCGCGCAGCCGCGCCAGCCTGGCTGACCTGACCCGCCCCGACCCGCGGCGG  790

Qy     67  TGCCCCAGATGCTGCCCCAGCCGCCAGCCGCGCTGCTTCTTCATCCTGGACGGCGCCGGACGAGC  126
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

```
source          1..6461
                /organism="Homo sapiens"
                /mol_type="genomic DNA"
                /db_xref="taxon:9606"
CDS             1..5556
                /note="unnamed protein product"
                /codon_start=1
                /protein_id="CAC34689.1"
                /db_xref="GI:13397120"
                /translation="MDDPEAPCSSTGPRLAVARELLAALEELSQEQLKRFRHKLRDV
                GPDGRSIPWGRLERADAVDLAEQLAQFYGPEPALEVARKTLKRADARDVAAQLQERRL
                QRIGLGSGTLLSVSEYYKKYREHVLQHARVKERNARSVKITKRFTKLLIAPESAAPE
                EALGPAEEPEPGRARRSDTHTFNRLFRRDEEGRPLTVVLQGPAGIGKTMAAKKILYD
                WAAGKLYQGQVDFAFFMPCGELLERPGTRSLADILDQCPDRGAPVPQMLAQPQRLLF
                ILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLVTTRAAAPGRL
                QGRLCSPQCAEVRGFSDKDKKKYFYKFPFRDRRAERAYRFVKENETLFALCFVPFVCW
                IVCTVLRQQLELGRDLSRTSKTTTSVLLFITTSVLSSAPVADGFRLQGDLRNLCRLAR
                EGVLGRRAQFAEKELEQLERGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFLAA
                LSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLGLLSAERMRDIRHFGCM
                VSERVKQEALRWVQGQGGCPGVAPEVTEGAKGLEDTEEPEEEEGEPNYPLELLYC
                LYETQBDAFVRQALCRFPELAIQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAA
                QEKKKSLGKRLQASLGGSWLGTQLAPEVPRPPCCDICPTPPDPDLLQSLSLSPTPSSF
                PLNIAPIQPLRRGLASVERMNVTVLAGAGPDPKTHAMTDPLCHLSSLTLSHCKLPDA
                VCRDLSEALRAAPALITELGLLHNRLSEAGLRMLSEGLAWPQCVQTVRVQLPDPQRGL
                QYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHOGCGLQTLSLSLSPSTPSSF
                SGRCREPGRRLGLESRWPRSAPEPSGDSEARTQVEAAGGAGGGRRRGREPPARGPHPQ
                PPRDAARGPGSSFAHSGRFVQGTPGPRTRPLPAGTEGSRGRESTSRPRARPSD
                RPRRPGTAPASQRPPGPGSRGPRTFLVARPQGGSGFLPALAWSRGTQVPTLAPGDRVG
                GDTHRAMLPSPLNVRLEAPAGMGEQLTETFALDTNTGSFLHTLEGPRFPRASQCIYAH
                RTIPHVLAFRVSIARLAPGSGPITLLLRSAFSPESPDLDLHGGPDFQGARYLYGHTLT
                PEQPGGPPQQEVHMLWTPAPPDLTLGGEEARTWDFLTAVGSQAEAQACLTEALQLQA
                RGALYTAHAQAWAQLWEVCGLDDVGFPLQLRQALRGSLYYLLSALPQPAPGYICHGJS
                PGGLSNGSREECYWGHVFWDDDLWMFPSILMFHPEAARAILEYRIRTLDGALENAQNL
                GVVGAKPAWESADSGLEVCPEDIYGVQEHVHNGAVVLAFELYYHTTQDLQLFREAGGK
                DVVRAVAEFWCSRVEWSPREEKYHLRGVMSPDEYHSGVNNSVTNVLVQNSLRFAAAL
                AQDLGLP1PSQWWAVADKIKVPFDVEQNFHPEFDGYEPDPRVCGTPSQRHLPVGEV
                VKQADVVLLGYPVPFLSLSPDVRRKNLEIYEAVTSPQGPAMTWKSMFAVGWMELKDAVRA
                RGLLDRSFANMAEPFKVWTENADGSGAVNFLTCMGGFLQAVVFGCTGFRVTRAGVTFD
                PVCLSGISRVSVSGIFYQKNLNFSFEDSVTVEVTARAGPWAPHLEAELWPQSRLS
                LPGHKVSFPRSAGRIQMSPPKLPGSSSSEPPGRTFSDVRDFLQSPLWVTLGSSSPTE
                SFTVDPASE"
```

```
BASE COUNT      1079 a    2146 c    2150 g    1086 t
ORIGIN
```

```
Qy     Match
Best Local Similarity   58.4%;       Score 2571.2;  DB 6;   Length 6461;
Matches 1772; Conservative  89.6%;   Pred. No. 4.5e-236;
                           Matches 1772;  Mismatches  0;  Indels 38;  Indels 168;  Gaps  :;
```

```
Qy        7  CGTGGGCGCGGCCAGCCTGCTGACCTGACCAGTGCCCGGCGCCCGGACCCGGCGCCCGG   66
              |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||
Db      731  CGGGGACAGCGGCCAGCCGGCCGGCTGATCCTGA CCAGTGCCCCGGCGCCCCGGCGCCGG  790

Qy       67  TGCCCCAGATGCTGCGCCCAGCCCGCAGCGGCCGCGCCCCTTCATCCTGACGCGGCGAGC  126
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      791  TGCCGCAGATGCTGCGGCCCAGCCCGCAGCGGCCGCGCCTGCTTCATCCTGACGCGGCGAGC  850

Qy      127  TGCCGGCTGGGCGGGCGCCCCGGGGGCCCCGGAGGCCGGCCGCGCCCTTGAGGCGGCGAGCG  186
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      851  TGCCGGCTGGGCGGGCGCGGCGCCCCGGGAGGCCGGCGCCCTT CGAGGCGGCGAGCG  910

Qy      187  GGCGGCGGTGCTAGGCGGGGCTGCTGAGTAAGCGCTCCACGGCCCCTGCTGG  246
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      911  GCGCGGCGGTGCTAGGCGGGGCTGCTGAGGAGCGCGTGCTCCCACGGCGCCTGCTGG  970

Qy      247  TGACCACGGCGGCGCCCCCCGGCCTTGTGTTCCCGGACGTGCG  306
              |||||||||||||||||||||||||||||||||||||||||
Db      971  TGACCACGGCGGCGCCGCCGGAGGGCTGCAGGGGCTGCAGGGGCTGCAGGGGCTGCG  1030

Qy      307  CCGAGGTGCGCGGCTTCTCGCAGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTCCGGG  366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     1031  CCGAGGTGCGCGGCTTCTCGCAGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTCCGGG  1090
```

```
Db    1871  GCCTGTAGGAGGACGCGGAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGC  1930
Qy    1207  TGGGCGCTGCAGGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGA  1266
Db    1931  TGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCGTGTTCTGAGCTACTGCGTGA  1990
Qy    1267  GGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCGAGATTGGTTGCTGCGCAGG  1326
Db    1991  GGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGA-TTGGTTGCTGCGCAGG  2050
Qy    1327  AGAGAGAAGAAGAGAGGCTGGGGAAGCAGCGGCTCCAGGCCAGCCTGGTGGCGGGCAGTTCTC  1386
Db    2051  AGAGAGAAGAAGAGAGGCTGGGGAAGCGGCTCCAGCCCAGCCAGCCGGTGGGCGGCAGCTGGC  2110
Qy    1387  AAGGCACCACAAAACAACTGGCCAGCCTCCCCTTCTTCATCCACTCTTT----------  1433
Db    2111  TGGGGACCCAACTGGCTCCAGAGTACCCTTTCGACCACCCTGCTGTGCCCATCGCCCCCA  2170
Qy    1434  ------------------------------------------------------  1433
Db    2171  CACCTCCACCAGACCCCTCGCCCCTCCCAGGGCAAGGCTTTTGCCAGAGTCCCTTGAATA  2230
Qy    1434  ------------------------------------------------------  1433
Db    2231  TAGCTCCAATTCAGCCCCTGCCCAGGGGCTTGGCATCTGTTGAGAGGATGAAGTCACGG  2290
Qy    1434  --------------------------------CAGGCAATGACTGACCCACTGTGCC  1458
Db    2291  TGTTGCAGGGCTGGCCTCACCCGCTGCCCCACTGCAAAC-CCCTGACGCGGTCTGCCGAGACCTT  2350
Qy    1459  ATCTGAGCAGCCTCACGCGCTGCCCCACTGCAAACTCCCTGACGCGGTCTGCCGAGACCTTT  1518
Db    2351  ATCTGAGCAGCCTCACGCGCTGCCCCACTGCAAACTCCCTGACGCGGTCTGCCGAGACCTTT  2410
Qy    1519  CTGAGGCCCTGAGGGCAGCCCCCGACTGACGGAGCTGGGCCTCCTCCACAACAGGCTCA  1578
Db    2411  CTGAGGCCCTGAGGGCAGCCCCCGACTGACGGAGCTGGGCC-CCTCCACAACAGGCTCA  2470
Qy    1579  GTGAGGCGGGACTGCCTATGCTGAGGGAGGGGCTAGCCTGGCCCCGGCAGTGCAGGGTGCAGA  1638
Db    2471  GTGAGGCAGCAGACTGCCTATGCTGAGGGAGGGGCCTAGCCTGCGCCCAGCCCAGGGGTGCAGA  2530
Qy    1639  CGGTCAGGGTACAGTCGCCTGACCCCGAGGGCC-CCAGTACCTGGTAGGGTATGCTTC  1698
Db    2531  CGGTCAGGGTACAGTCGCCTGACCCCGAGGGCC-CCAGTACCTGGTAGGGTATGCTTC  2590
Qy    1699  GGCAGAGCCCGGCCCTGACCACCCCTGGATCTCAGCGGGCTCAGCCGGCTGCCAACTGCCCCCCATGG  1758
Db    2591  GGCAGAGCCCGGCCCTGACCACCCCTGGATCTCAGCGGGCTCAGCCGGCTGCCAACTGCCCCCCATGG  2650
Qy    1759  TGACCTACCTGTGTGCAGTCTGCGCACCAGGGATGCGGGCTGCGGCCGAGACCCTGCAGACCTT  1816
Db    2651  TGACCTACCTGTGTGCAGTCTGCGCACCAGGGATGCGGGCTGCGGCCGAGACCCTGCAGCCT  2705
```

```
RESULT 9
AX088208
LOCUS       AX088208
DEFINITION  Sequence 3 from Patent WO0114564.
ACCESSION   AX088208
VERSION     AX088208.1  GI:13397119
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Peyman,J.A., Green,C.D., Hsu,A., Browning,S.A. and Carulli,J.
  TITLE     Polynucleotides expressed in activated t-lymphocytes and proteins
            encoded thereby
  JOURNAL   Patent: WO 0114564-A 3 01-MAR-2001;
            Curagen Corporation (US) ; BIOGEN, INC. (US)
FEATURES    Location/Qualifiers
```

RESULT 10
AF015416
LOCUS        AF015416          95038 bp    DNA         linear   PRI 20-AUG-1997
DEFINITION   Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
ACCESSION    AF015416
VERSION      AF015416.1  GI:2335202
KEYWORDS     HTG.
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 95038)
  AUTHORS    Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
             Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
             Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
             Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
             Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.
             Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
             Wilson,R. and Burbee,D.
  TITLE      95 Kb from Four Overlapping Human Chromosome 11p15.5 Cosmids
  JOURNAL    Unpublished
REFERENCE    2  (bases 1 to 95038)
  AUTHORS    Evans,G.A., Bradbury,P., Brignac,S., Bumeister,R., Davie,J.,
             Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
             Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,J.,
             Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
             Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.
             Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
             Wilson,R. and Burbee,D.
  TITLE      Direct Submission
  JOURNAL    Submitted (18-JUL-1997) Genome Science and Technology Center,
             University of Texas Southwestern Medical Center atDallas, 5323
             Harry Hines Blvd, Dallas, TX 75235-8591, USA
FEATURES             Location/Qualifiers
     source          1..95038
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"

```
                /db_xref="taxon:9606"
                /chromosome="11"
                /map="11p15.5"
repeat_region   711..975
                /rpt_family="Alu"
repeat_region   2728..3022
                /rpt_family="Alu"
repeat_region   3363..3719
                /rpt_family="THE1"
repeat_region   5634..5909
                /rpt_family="Alu"
repeat_region   complement(6720..6959)
                /rpt_family="Alu"
repeat_region   10560..10839
                /rpt_family="Alu"
repeat_region   complement(11832..12455)
                /rpt_family="Alu"
repeat_region   12921..13405
                /rpt_family="L1"
repeat_region   13718..14314
                /rpt_family="Alu"
repeat_region   14617..14779
                /rpt_family="Alu"
repeat_region   15336..15695
                /rpt_family="Alu"
repeat_region   15722..15994
                /rpt_family="Alu"
repeat_region   complement(16696..16997)
                /rpt_family="Alu"
repeat_region   complement(18996..19286)
                /rpt_family="Alu"
repeat_region   complement(19458..19747)
                /rpt_family="Alu"
repeat_region   complement(19777..20045)
                /rpt_family="Alu"
repeat_region   complement(22885..23002)
                /rpt_family="MIR"
repeat_region   23109..23261
                /rpt_family="Alu"
repeat_region   23292..23481
                /rpt_family="MER46"
repeat_region   complement(23877..23944)
                /rpt_family="MIR"
repeat_region   complement(24695..24986)
                /rpt_family="Alu"
repeat_region   complement(25429..26012)
                /rpt_family="Alu"
repeat_region   complement(27380..27676)
                /rpt_family="Alu"
repeat_region   complement(29422..29486)
                /rpt_family="MER42"
repeat_region   complement(29514..29802)
                /rpt_family="Alu"
repeat_region   30797..31075
                /rpt_family="Alu"
repeat_region   31688..31792
                /rpt_family="MLT2B2"
repeat_region   32125..32412
                /rpt_family="Alu"
repeat_region   complement(38871..39186)
                /rpt_family="Alu"
repeat_region   42952..43094
                /rpt_family="MER5"
repeat_region   43106..43388
                /rpt_family="MER42"
repeat_region   43754..44216
                /rpt_family="Alu"
repeat_region   complement(44815..45136)
                /rpt_family="MER42"
repeat_region   45145..45484
                /rpt_family="Alu"
repeat_region   45740..45995
                /rpt_family="Alu"

repeat_region   46375..46646
                /rpt_family="Alu"
repeat_region   47145..47738
                /rpt_family="Alu"
repeat_region   48353..48626
                /rpt_family="Alu"
repeat_region   49567..49842
                /rpt_family="Alu"
repeat_region   49940..50211
                /rpt_family="Alu"
repeat_region   complement(50615..50907)
                /rpt_family="Alu"
repeat_region   complement(53001..53290)
                /rpt_family="Alu"
repeat_region   complement(54161..54512)
                /rpt_family="Alu"
repeat_region   54533..54822
                /rpt_family="L1"
repeat_region   57899..58113
                /rpt_family="Alu"
repeat_region   complement(58795..59044)
                /rpt_family="Alu"
repeat_region   complement(59414..60023)
                /rpt_family="L1"
repeat_region   complement(60177..60311)
                /rpt_family="Alu"
repeat_region   complement(60770..61108)
                /rpt_family="L1"
repeat_region   complement(61207..61557)
                /rpt_family="THE1"
repeat_region   61670..61942
                /rpt_family="Alu"
repeat_region   61944..62436
                /rpt_family="MER41"
repeat_region   complement(67807..68105)
                /rpt_family="Alu"
repeat_region   70961..71049
                /rpt_family="Alu"
repeat_region   71224..71523
                /rpt_family="Alu"
repeat_region   complement(85033..85397)
                /rpt_family="MER4"
repeat_region   complement(85414..85620)
                /rpt_family="Alu"
repeat_region   complement(85701..85804)
                /rpt_family="MER4"
repeat_region   complement(85895..86028)
                /rpt_family="MER4"
repeat_region   86180..86477
                /rpt_family="Alu"
repeat_region   86497..86780
                /rpt_family="Alu"
repeat_region   87125..87256
                /rpt_family="Alu"
repeat_region   87590..87893
                /rpt_family="Alu"
repeat_region   89103..89401
                /rpt_family="MER4"
repeat_region   89726..89837
                /rpt_family="MER4"
repeat_region   89888..90390
                /rpt_family="MER4"
repeat_region   complement(91891..92187)
                /rpt_family="Alu"
repeat_region   complement(93951..94017)
                /rpt_family="LTR12"
BASE COUNT      23552 a    24227 c    24425 g    22834 t
ORIGIN

Query Match              50.9%;  Score 1369;  DB 9;  Length 95038;
Best Local Similarity    99.6%;  Pred. No. 1.2e-204;
Matches 1372;  Conservative 0;  Mismatches 5;  Indels 0;  Gaps 0;
```

```
Qy       7  CGTGGGCGCGGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCGCGGCCGACCGGGCGCCGG     66
               ||||||||||||||||| ||||||||||| ||||||||| |||||||||||||||||||||
Db   64958  CGGGGCACGCGGCAGCCTGGC-GACCTGATCCTGGACCAG-GCCCGCGGCCGACCGGGCGCCGG  65017

Qy      67  TGCCCAGATGCTGGCCCAGCCGCGGCGCGTCTTCATCTTGGACAGACCCCTTCGAGGGCGGAGCG    126
               ||||||||||||| ||||||||||||| ||||||||||| ||||||  |||||||||||||||
Db   65018  TGCCCAGATGCTGGCGCAGCCGCGGCGCGTCTTCATCTTGGACAGACCCCTTCGAGGGCGGAGCG  65077

Qy     127  TGCCGGCGCTGGGGGGCCCCAGGCGCGGCGCCCTGCACAGACCCCTTCGAGGGCGGAGCG       186
Db   65078  TGCCGGCGCTGGGGGGCCCCAGGCGCGGCGCCCTGCACAGACCCCTTCGAGGGCGGAGCG     65137

Qy     187  GCGCGCGGGTGCTAGGCGGGTGCTGAGCAAGGCGCTGCTGAGCAAGGCGCCTCCTGCTGG      246
Db   65138  GCGCGCGGGTGCTAGGCGGGTGCTGAGCAAGGCGCTGCTGAGCAAGGCGCCTCCTGCTGG    65197

Qy     247  TGACCACGCGCGCGCCCGCCCCGGGAGGCTGCAGGGCGCTGTTCCCCGCAGTGCGG        306
Db   65198  TGACCACGCGCGCGCCCGCCCCGGGAGGCTGCAGGGCGCTGTTCCCCGCAGTGCGG      65257

Qy     307  CCGAGGTGCCGGCTTCTCTCCGACAAGGACAAGAAGAAGTATTCTACACAAGTTCTTCCCGGG       366
Db   65258  CCGAGGTGCCGGCTTCTCTCCGACAAGGACAAGAAGAAGTATTCTACACAGTTCTTCCCGGG     65317

Qy     367  ATGAGAGGAGGGCCAGCCGCGCGCCCCACCGCGCGCCACCGCGCGCCGCGTGTCGCC        426
Db   65318  ATGAGAGGAGGGCCAGCCGCGCGCCCCACCGCGCGCCACCGCGCGCCGCGTGTCGCC      65377

Qy     427  TGTGCTTCGTGCCCTTCGTGGCTGGATCGTGTGCACCGCTGCTGCCGCAGCTGGAGC        486
Db   65378  TGTGCTTCGTGCCCTTCGTGGCTGGATCGTGTGCACCGCTGCTGCCGCAGCTGGAGC      65437

Qy     487  TCGGTCGGGACCTGCTGCGGCCACGTCCAAGACCTCAGTGTACCTGCTTTCATCCA        546
Db   65438  TCGGTCGGGACCTGCTGCGGCCACGTCCAAGACCTCAGTGTACCTGCTTTCATCCA      65497

Qy     547  CCAGGCTTCTGAGCTCGGCTTCGGTGGCTCGGTTGCAGGGCGACCTGCGCGCA        606
Db   65498  CCAGGCTTCTGAGCTCGGCTTCGGTGGCTCGGTTGCAGGGCGACCTGCGCCA      65557

Qy     607  ATCTGTGCCGCCTGCCCGCGAGGGCGTCCTCGGACGCAGGGGCAGTTGCCGAGAAGG        666
Db   65558  ATCTGTGCCGCCTGCCCGCGAGGGCGTCCTCGGACGCAGGGGCAGTTGCCGAGAAGG      65617

Qy     667  AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCCTGTTCTCAGCAAAAAGG       726
Db   65618  AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCCTGTTTCTCAGCAGAGAAGG   65677

Qy     727  AGCTGCCCGGCGTGCTGGAGCACAGAGGTCACCAGTTCATCCGACCAGAGCTTCCAGG       786
Db   65678  AGCTGCCCGGCGTGCTGGAGCACAGAGGTCACCAGTTCATCCGACCAGAGCTTCCAGG     65737

Qy     787  AGTTCCCGCGGCACTGTCCTACCTGCTCGGGGGTGCCCAGGACCCGCGGCTG        846
Db   65738  AGTTCCCGCGGCACTGTCCTACCTGCTCGGGGGTGCCCAGGACCGCGGCTG      65797

Qy     847  GCGGCGTTGGGACACTCCTGCTGCGTGGGACGCCCAGCCGCCACAGCCACTGGTGCTCACCA       906
Db   65798  GCGGGCGTTGGGACACTCCTGCTGCGTGGGACGCCCAGCCGCCACTGGTGCTGCTCACCA     65857

Qy     907  CGCGCTTCCTCTTCGGACTGCTCGGACTGCTGGAGCGGCGGAGCCGGACATGCGCGACATCGAGCGCCACTTCG       966
Db   65858  CGCGCTTCCTCTTCGGACTGCTCGGACTGCTGGAGCGCGGAGCGGGACATGCGCGACATCGAGCGCCACTTCG     65917

Qy     967  GCTCATGGTTCAGAGACGTGGCCGTGGTGCAGGGACAGGGAC       1026
Db   65918  GCTCATGGTTTCAGAGACGTGGAGCGTGGGTGCGAGGGGACAGGGAC    65977

Qy    1027  AGGGCTGCCCGGAGTGCCGGCCACCAGAGTGACCCAGAGGGCTCGAGGACACCG      1086
Db   65978  AGGGCTGCCCCGGAGTGCCGGCCACCAGAGTGACCCAGAAAGGGCTCGAGGACACCG    66037

Qy    1087  AAGAGCCAGAGGAGGAGGAGGAGGAGGAGGCCAACTACCCACTGAGTTGCTGACT       1146
Db   66038  AAGAGCCAGAGGAGGAGGAGGAGGAGGAGGAGGCCCAACTACCCACTGAGTTGCTGACT     66097

Qy    1147  GCCTGTACGAGACGACGCAGGAGGACGCGGTTGTGCGCGCCAAGCCCTGTGCGCGGTTCCCGGAGC   1206
Db   66098  GCCTGTACGAGACGACGCAGGAGGACGCGGTTGTGCGCGCCAAGCCCTGTGCGCGGTTCCCGGAGC   66157

Qy    1207  TGGCGCTGCAGCCAGTGCGCTTCTGCGCCATCGACGTGGCGTCTTCTGAGCTACTGCCTGA   1266
Db   66158  TGGCGCTGCAGCCAGTGCGCTTCTGCGCCATCGACGTGGCGTCTTCGAGCTACTGCCTGA   66217

Qy    1267  GGTGCTGCCCTGCTGGACAGGCACTGCGGCGTGATCAGCTGCAGATTGGTTCTGCCGCAGG   1326
Db   66218  GGTGCTGCCCTGCTGGACAGGCACTGCGGCGTGATCAGCTGCAGATTGGTTCTGCCGCAGG   66277

Qy    1327  AGAAGAAGAAGAGAGCCTGGGAAGCGGCTCCAGGCAGCCAGCCTGGGTGCGGCCAGT    1383
Db   66278  AGAAGAAGAAGAGAGCCTGGGAAGCGGCTCCAGGCAGCCAGCCTGGGTGCGGCCAGT   66334
```

RESULT 11
AC136475
LOCUS       AC136475       143835 bp    DNA       linear   PRI 30-MAY-2003
DEFINITION  Homo sapiens chromosome 11, clone RP11-326C3, complete sequence.
ACCESSION   AC136475
VERSION     AC136475.6  GI:31193978
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-326C3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,
            Barna,N., Bastien,V., Bloom,T., Boguslavkiy,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagoplan,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Lui,A., Mabbitt,R., MacLean,C.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,

```
          Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
          Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
          Talamas,J., Tesfaye.S., Theodore,J., Topham,K., Travers,M.,
          Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   4  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
          Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
          Boguslavky,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
          Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
          Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
          Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
          Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
          Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
          Karat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
          Lindblad-Toh,K., Liu,C., Lui,A., Mabbitt,R., MacLean,C.,
          Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
          Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
          Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
          O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
          Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
          Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
          Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
          Talamas,C., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
          Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
          Wyman,D., Young,G., Zainoun,J., Zembek,J., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (30-MAY-2003) Whitehead Institute/MIT Center for Genome
          Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT   On May 30, 2003 this sequence version replaced gi:29469599.
          All repeats were identified using RepeatMasker.
          Smit, A.F.A. & Green, P. (1996-1997)
          http://ftp.genome.washington.edu/RM/RepeatMasker.html
          ----------- Genome Center
            Center: Whitehead Institute/ MIT Center for Genome Research
            Center code: WIBR
            Web site: http://www-seq.wi.mit.edu
            Contact: sequence_submissions@genome.wi.mit.edu
          ----------- Project Information:
            Center project name: L28535
            Center clone name: 326_C_3
          -----------

          Only the first 143.8 kilobases of this clone are being submitted.
          The remainder overlaps accession number AC138230 [WICGR project
          L29013].
FEATURES             Location/Qualifiers
  source          1..143835
                  /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
                  /chromosome="11"
                  /map="11"
                  /clone="RP11-326C3"
                  /clone_lib="RPCI-11 Human Male BAC"
  repeat_region   1712..2006
                  /rpt_family="AluSq"
  repeat_region   2347..2703
                  /rpt_family="THE1B"
  repeat_region   4157..4302
                  /rpt_family="L1ME4a"
  repeat_region   4435..4455
                  /rpt_family="AT_rich"
  repeat_region   4528..4591
                  /rpt_family="AT_rich"
  repeat_region   4592..4901
                  /rpt_family="AluSx"
  repeat_region   4902..4968
                  /rpt_family="AluSx"
  repeat_region   4979..5091
                  /rpt_family="AT_rich"
```

```
  repeat_region   5154..5251
                  /rpt_family="L1ME4a"
  repeat_region   5620..5683
                  /rpt_family="L1ME4a"
  repeat_region   complement(5684..5954)
                  /rpt_family="L1ME4a"
  repeat_region   5955..6050
                  /rpt_family="AluJo"
  repeat_region   complement(6059..6206)
                  /rpt_family="L1ME4a"
  repeat_region   6349..6578
                  /rpt_family="MER46C"
  repeat_region   6792..7203
                  /rpt_family="L1ME4a"
  repeat_region   complement(8884..9052)
                  /rpt_family="L2"
  repeat_region   9225..9287
                  /rpt_family="L1MC5"
  repeat_region   9676..9986
                  /rpt_family="MER68-int"
  repeat_region   9994..10032
                  /rpt_family="AluSg"
  repeat_region   10182..10417
                  /rpt_family="AT_rich"
  repeat_region   complement(10418..10946)
                  /rpt_family="L1MEd"
  repeat_region   complement(10947..11251)
                  /rpt_family="LTR59"
  repeat_region   complement(11252..11272)
                  /rpt_family="AluSx"
  repeat_region   complement(11290..11595)
                  /rpt_family="LTR59"
  repeat_region   11702..12014
                  /rpt_family="AluSp"
  repeat_region   12047..12137
                  /rpt_family="AluSg"
  repeat_region   complement(12159..12465)
                  /rpt_family="L1M4"
  repeat_region   12466..12605
                  /rpt_family="AluSx"
  repeat_region   12763..12832
                  /rpt_family="L2"
  repeat_region   12898..13715
                  /rpt_family="L2"
  repeat_region   13716..14007
                  /rpt_family="L1MB5"
  repeat_region   14018..14320
                  /rpt_family="AluSc"
  repeat_region   complement(16654..16949)
                  /rpt_family="AluJo"
  repeat_region   17027..17095
                  /rpt_family="AluSx"
  repeat_region   14321..14603
                  /rpt_family="L1MC5"
  repeat_region   14611..14772
                  /rpt_family="L1MB5"
  repeat_region   15386..15685
                  /rpt_family="AluSg/x"
  repeat_region   15701..15984
                  /rpt_family="AluSx"
  repeat_region   19299..19409
                  /rpt_family="AluY"
  repeat_region   complement(19426..19719)
                  /rpt_family="L2"
  repeat_region   complement(19742..20040)
                  /rpt_family="AluSg"
  repeat_region   20207..20435
                  /rpt_family="AluJo"
  repeat_region   20906..21032
                  /rpt_family="L1MC/D"
  repeat_region   complement(18948..19258)
                  /rpt_family="(CCCCG)n"
```

repeat_region    complement(21181..21341)
                 /rpt_family"MIR"
repeat_region    complement(22663..22828)

Query Match     50.8%;  Score 1365.8;  DB 9;  Length 143835;
Best Local Similarity 99.5%;  Pred. No. 3.5e-204;
Matches 1370;  Conservative 0;  Mismatches 0;  Indels 0;  Gaps 0;

```
Qy       7 CGTGGGCGGCCAGCCCGGCTGACCTGATCCTGGACCAGTGCCCGACCCGGCGGGGCGGCGG 66
             ||||||||||||||||| ||||||||||||||||||||||||  ||||||||||||||||||
Db   65009 CGGGCACGCGGCCAGCCCGGCTGACCTGATCCTGGACCAGTGCCCGACGCGGCGGGGCGCCGG 65068

Qy      67 TGCCCGAGATGCTGGCCAGGTGCCAGGCGGCTGCTCTTCATCTCTGGAGACGGCGGCGACGAGC 126
             | |||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db   65069 TGCCCGAGATGCTGGCCAGGTGCCAGGCGGCTGCTCTTCATCTCTGGAGACGGCGGCGACGAGC 65128

Qy     127 TGCCCGCGTGGGGGCGCCCGAGGCCGCGGCCCCTTCGCACAGACCCCTTCGAGGCGGCGACGCG 186
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65129 TGCCCGCGTGGGGGCGCCCGAGGCCGCGGCCCCTTCGCACAGACCCCTTCGAGGCGGCGACGCG 65188

Qy     187 GCGGCGGGGTGCTAGGCGGGCTGCGAGTAAGGCGGCTGCTGCCCACGGCCCTCCTGGTGGG 246
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65189 GCGCGCGGGTGCTAGGCGGGCTGCGAGTAAGGCGGCTGCTGCCCACGGCCCTCCTGGTGGG 65248

Qy     247 TGACCACGGCGGGGCGCGCGCCGCCCGGGGAGGCTGCAGGGCGCCCGTGTTCCCCGCAGTGCGG 306
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   55249 TGACCACGGCGGGGCGCGCGCCGCCCGGGGAGGCTGCAGGGCGCCCGTGTTCCCCGCAGTGCGG 65308

Qy     307 CCGAGGCGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 366
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65309 CCGAGGCGCGGCGGCTTCTCCCACAAGGACAAGAAGAAGTATTTCTACAAGTA--TCCCGGG 65368

Qy     367 ATGAGGAGGGCCGAGCGCGGCCCTACCGCGCTTCGTGAAGGAGAACGAGACGCGTCTCGCGC 426
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65369 ATGAGGAGGGCCGAGCGCGGCCCTACCGCGCTTCGTGAAGGAGAACGAGACGCGTCTCGCGC 65428

Qy     427 TGTGCTTCGTGCCCTTCGTGCTGATCGTGTGTGCACCGTGCGCGCAGCTGGAGC 486
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65429 TGTGCTTCGTGCCCTTCGTGCTGATCGTGTGTGCACCGTGCGCGCAGCTGGAGC 65488

Qy     487 TCGGTCGGGACCGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTCATCA 546
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65489 TCGGTCGGGACCGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTTTTCATCA 65548

Qy     547 CCAGCGTTCTGAGCGCGGCTCGGCTCGGTAGCCGCACGGGCCCCGGTTCGAGGGGCCACCTGCGCA 606
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65549 CCAGCGTTCTGAGCGCGGCTCGGCTCGGTAGCCGCACGGGCCCCGGTTCGAGGGGCCACCTGCGCCA 65608

Qy     607 ATCCGTGCCCGCTGGCCCCTGCGAGGGCGTCCTCGACGCAGTGGCCCCAGT-TGCCGAGAAGG 666
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65609 ATCTGTGCCCGCTGGCCCCGGAGGGCGTCCTCGGACGCCAGGGCCAGT-TGCCGAGAAGG 65668

Qy     667 AACTGGAGCAACTGGAGCTTCGTGTCGCTGCTCCAAAGTGCACTCTGTTTCTCAGCAAAAAGG 726
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65669 AACTGGAGCAACTGGAGCTTCGTGTCGCTGCTCCAAAGTGCACTCTGTTTCTCAGCAAAAAGG 65728

Qy     727 AGTCTGCCGGCGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 786
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65729 AGTCTGCCGGCGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 65788

Qy     787 AGTTCCTGCGCGCACTGTCCTACCTGCTGCTGGAGACGGCGGGGGTGCCCAGGACGCGGGCTG 846
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65789 AGTTCCTGCGCGCACTGTCCTACCTGCTGCTGGAGACGGCGGGGGTGCCCAGGACGCGGGCTG 65848

Qy     847 GCGGGCTTGGGACACTCCTGCGTGGGACGCCCAGCGCCGCACAGCCCACTGGTGTGC-CACCA 906
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65849 GCGGGCTTGGGACACTCCTGCGTGGGACGCCCAGCGCCGCACAGCCCACTGGTGTGC-CACCA 65908

Qy     907 CGCGCTTCTCTTCCGGACTGCTGAGCGCGGAGCGATGCGCGACATCGAGCGGCCACTTCG 966
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   65909 CGCGCTTCTCTTCCGGACTGCTGAGCGCGGAGCGATGCGCGACATCGAGCGGCCACTTCG 65968

Qy     967 GCTGCATGGTTTCAGACGGTGTGAAGCAGCAGGAGCCCTGCGGTGGTGCGGGGACAGGGAC 1026

Db   65969 GCTGCATGGTCTTCAGACGGTGCCACCAGAGTGCAAGCAGGAGGAGGACCAGGACCCGAGGACACCG 66028

Qy    1027 AGGGCTGCCCCGAGTGCCACCAGAGTGCCACCAGAGTGACCAGAGGGCCAAAGGCTCGAGGACACCG 1086
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   56029 AGGGCTGCCCCGAGTGCCACCAGAGTGCCACCAGAGTGACCAGAGGGCCAAAGGCTCGAGGACACCG 66088

Qy    1087 AAGAGCCAGAGGAGGAGGAGGAGGAGGAGCCCAACTACCCACTGGAGTTGCTGTACT 1146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66089 AAGAGCCAGAGGAGGAGGAGGAGGAGGAGCCCAACTACCCACTGGAGTTGCTGTACT 66148

Qy    1147 GCCTGTACGAGACGCCAGGAGACGCGTTGTCGCCCAAGGCCTGTGCCGAGC 1206
             ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66149 GCCTGTACGAGACGCCAGGAGACGCGTTGTCGCCCAAGGCCTGTGCCGAGC 66208

Qy    1207 TGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGATGCGCATGGAGCTACTGCGTGA 1266
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66209 TGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGATGCGCATGGAGCTACTGCGTGA 66268

Qy    1267 GGTGCTGCCCTGCTGGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGG 1326
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66269 GGTGCTGCCCTGCTGGGACAGGCACTGCGGCTGATCAGCTGCAGAT-GGTTGCTGCGCAGG 66328

Qy    1327 AGAAGAAGAAGAAGAGCCTGGGGAAGCCTGGGGAACGCGCTCCAGCCACCTGGGGTGGCGCCAGTT 1383
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   66329 AGAAGAAGAAGAAGAGCCTGGGGAAGCCTGGGGAACGCGCTCCAGCCACCTGGGGTGGCGCCAGGT 66385
```

RESULT 12
AP006283

LOCUS       AP006283              156819 bp    DNA     linear   HTG 26-MAR-2003
DEFINITION  Homo sapiens chromosome 11 clone RP11-326C3 map 11p, WORKING DRAFT
            SEQUENCE, 19 unordered pieces.
ACCESSION   AP006283
VERSION     AP006283.1  GI:29243343
KEYWORDS    HTG; HTGS PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA of 11p
  JOURNAL   Published Only in Database (2003)
REFERENCE   2  (bases 1 to 156819)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     -------------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: R-KEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            -------------- Project Information
            Center project name: HumDraft11
            Center clone name: RP11-326C3
            -------------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 144706 bases at least Q40
            Consensus quality: 150967 bases at least Q30
            Consensus quality: 153657 bases at least Q20
            Insert size: 155019; sum-of-contigs
            Quality coverage: 6.36x in Q20 bases; sum-of-contigs
            --------------
            NOTE: This is a 'working draft' sequence. It currently consists of

19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved

```
         1      21001  contig of   21001 bp in length
     21102     40714  contig of   19613 bp in length
     40815     58430  contig of   17616 bp in length
     58531     66754  contig of    8224 bp in length
     66855     81134  contig of   14280 bp in length
     81235     89763  contig of    8529 bp in length
     89864    101889  contig of   12026 bp in length
    101990    112625  contig of   10636 bp in length
    112726    119802  contig of    7077 bp in length
    119903    125181  contig of    5279 bp in length
    125282    129184  contig of    3903 bp in length
    129285    137260  contig of    7976 bp in length
    137361    143542  contig of    6182 bp in length
    143643    146891  contig of    3249 bp in length
    146992    149568  contig of    2577 bp in length
    149669    152678  contig of    3010 bp in length
    152779    154135  contig of    1357 bp in length
    154236    155662  contig of    1427 bp in length
    155763    156819  contig of    1057 bp in length.
```

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

```
         1      21001: contig of 21001 bp in length
     21002     21101: gap of 100 bp
     21102     40714: contig of 19613 bp in length
     40715     40814: gap of 100 bp
     40815     58430: contig of 17616 bp in length
     58431     58530: gap of 100 bp
     58531     66754: contig of 8224 bp in length
     66755     66854: gap of 100 bp
     66855     81134: contig of 14280 bp in length
     81135     81234: gap of 100 bp
     81235     89763: contig of 8529 bp in length
     89764     89863: gap of 100 bp
     89864    101889: contig of 12026 bp in length
    101890    101989: gap of 100 bp
    101990    112625: contig of 10636 bp in length
    112626    112725: gap of 100 bp
    112726    119802: contig of 7077 bp in length
    119803    119902: gap of 100 bp
    119903    125181: contig of 5279 bp in length
    125182    125281: gap of 100 bp
    125282    129184: contig of 3903 bp in length
    129185    129284: gap of 100 bp
    129285    137260: contig of 7976 bp in length
    137261    137360: gap of 100 bp
    137361    143542: contig of 6182 bp in length
    143543    143642: gap of 100 bp
    143643    146891: contig of 3249 bp in length
    146892    146991: gap of 100 bp
    146992    149568: contig of 2577 bp in length
    149569    149668: gap of 100 bp
    149669    152678: contig of 3010 bp in length
    152679    152778: gap of 100 bp
    152779    154135: contig of 1357 bp in length
    154136    154235: gap of 100 bp
    154236    155662: contig of 1427 bp in length
    155663    155762: gap of 100 bp
    155763    156819: contig of 1057 bp in length.
```

```
FEATURES             Location/Qualifiers
     source          1..156819
                     /organism="Homo sapiens"
```

```
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p"
                     /clone="RP11-326C3"
     misc_feature    1..21001
                     /note="assembly_fragment"
     misc_feature    21102..40714
                     /note="assembly_fragment"
     misc_feature    40815..58430
                     /note="assembly_fragment"
     misc_feature    58531..66754
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    66855..81134
                     /note="assembly_fragment"
     misc_feature    81235..89763
                     /note="assembly_fragment"
     misc_feature    89864..101889
                     /note="assembly_fragment"
     misc_feature    101990..112625
                     /note="assembly_fragment"
     misc_feature    112726..119802
                     /note="assembly_fragment"
     misc_feature    119903..125181
                     /note="assembly_fragment"
     misc_feature    125282..129184
                     /note="assembly_fragment clone_end:T7 vector_side:right"
     misc_feature    129285..137260
                     /note="assembly_fragment"
     misc_feature    137361..143542
                     /note="assembly_fragment"
     misc_feature    143643..146891
                     /note="assembly_fragment"
     misc_feature    146992..149568
                     /note="assembly_fragment"
     misc_feature    149669..152678
                     /note="assembly_fragment"
     misc_feature    152779..154135
                     /note="assembly_fragment"
     misc_feature    154236..155662
                     /note="assembly_fragment"
     misc_feature    155763..156819
                     /note="assembly_fragment"
BASE COUNT    37089 a    39987 c    40186 g    37757 t     1800 others
ORIGIN
```

```
Query Match        50.8%;   Score 1365.8;   DB 2;   Length 156819;
Best Local Similarity   99.5%;   Pred. No. 3.5e-204;
Matches 1370;   Conservative   0;   Mismatches   7;   Indels   0;   Gaps   0;

Qy         7 CGTGGGGCCGCCAGCCTGGCTGGCTGCTGATCCTGGACCAGTGCCCGACCGGGCGGCGGCGCGCGG   66
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12579 CGGGCACGCCAGCCTGGCTGGCTGCTGATCCTGGACCAGTGCCCGACCGGGCGGCGGCGGACGAGC  12636

Qy        67 TGCCCAGATGCTGGCCCAGCCGCAGGCGCGGCTGCTTCATCCTGGACGCGCTGCTGACGCGACGAGC  126
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12639 TGCCCAGATGCTGGCCCAGCCGCAGGCGCGGCTGCTTCATCCTGGACGCGCTGCTGACGCGACGAGC  12698

Qy       127 TGCCGGCGCTGGGGGCCCGGGGCCCGGCCCGCCCGAGGCGCGGCTGCTGAGTAAGGCGCCTCGCTGG  186
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12699 TGCCGGCGCTGGGGGCCCGGGGCCCGGCCCGCCCGAGGCGCGGCTGCTGAGTAAGGCGCCTCGCTGG  12758

Qy       187 GGCGGCGGGGTGCTAGGCGGGGCTGCTGGCAAGGCGCCGCCCGCCTCCTGGCC  246
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12759 GCGCGGCGGGGTGCTAGGCGGGGCTGCTGGCAAGGCGCCGCCCGCCTCCTGCTGG  12818

Qy       247 TGACCACGGCCCGCGCGCCCCGGGAGGCTGCAGGGCGCCGCCCGCTGTTCCCGCAGTGCG  306
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12819 TGACCACGGCCCGCGCGCCCCCGGGAGGCTGCAGGGCGCCGCCCGCTGTTCCCGCAGTGCG  12878

Qy       307 CCGAGGTGCCGCGGCTTCTCCGACAAGGACAAGAGAAG:ATTTCTACAAGTTCTTCCGGG  366
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     12879 CCGAGGTGCCGCGGCTTCTCCGACACAAGGACAAGAGAAGAAGTTCTACAAGTATTTCCGGG  12938
```

DEFINITION  Mus musculus PYRIN-containing APAF1-like protein 5, mRNA (cDNA
clone MGC:37241 IMAGE:4972485), complete cds.
ACCESSION   BC031139
VERSION     BC031139.1  GI:21410401
KEYWORDS    MGC.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 3454)
AUTHORS     Strausberg,R.L., Feingold,E.A., Grouse,L.H., Derge,J.G.,
Klausner,R.D., Collins,F.S., Wagner,L., Shenmen,C.M., Schuler,G.D.,
Altschul,S.F., Zeeberg,B., Buetow,K.H., Schaefer,C.F., Bhat,N.K.,
Hopkins,R.F., Jordan,H., Moore,T., Max,S.I., Wang,J., Hsieh,F.,
Diatchenko,L., Marusina,K., Farmer,A.A., Rubin,G.M., Hong,L.,
Stapleton,M., Soares,M.B., Bonaldo,M.F., Casavant,T.L.,
Scheetz,T.E., Brownstein,M.J., Usdin,T.B., Toshiyuki,S.,
Carninci,P., Prange,C., Raha,S.S., Loquellano,N.A., Peters,G.J.,
Abramson,R.D., Mullahy,S.J., Bosak,S.A., McEwan,P.J.,
McKernan,K.J., Malek,J.A., Gunaratne,P.H., Richards,S.,
Worley,K.C., Hale,S., Garcia,A.M., Gay,L.J., Hulyk,S.W.,
Villalon,D.K., Muzny,D.M., Sodergren,E.J., Lu,X., Gibbs,R.A.,
Fahey,J., Helton,E., Ketteman,M., Madan,A., Rodrigues,S.,
Sanchez,A., Whiting,M., Madan,A., Young,A.C., Shevchenko,Y.,
Bouffard,G.G., Blakesley,R.W., Touchman,J.W., Green,E.D.,
Dickson,M.C., Rodriguez,A.C., Grimwood,J., Schmutz,J., Myers,R.M.,
Butterfield,Y.S., Krzywinski,M.I., Skalska,U., Smailus,D.E.,
Schnerch,A., Schein,J.E., Jones,S.J. and Marra,M.A.
TITLE       Generation and initial analysis of more than 15,000 full-length
human and mouse cDNA sequences
JOURNAL     Proc. Natl. Acad. Sci. U.S.A. 99 (26), 16899-16903 (2002)
MEDLINE     22288257
PUBMED      12477932
REFERENCE   2  (bases 1 to 3454)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (03-JUN-2002) National Institutes of Health, Mammalian
Gene Collection (MGC), Cancer Genomics Office, National Cancer
Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
Email: cgapbs-r@mail.nih.gov
Tissue Procurement: Jeffrey E. Green, M.D.
cDNA Library Preparation: Life Technologies, Inc.
cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
DNA Sequencing by: Sequencing Group at the Stanford Human Genome
Center, Stanford University School of Medicine, Stanford, CA 94305
Web site:       http://www-shgc.stanford.edu
Contact:   (Dickson, Mark) mcd@paxil.stanford.edu
Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
R. M.

Clone distribution: MGC clone distribution information can be found
through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
Series: IRAK Plate: 59 Row: c Column: 20
This clone was selected for full length sequencing because it
passed the following selection criteria: matched mRNA gi: 23956261.

FEATURES             Location/Qualifiers
source          1..3454
                /organism="Mus musculus"
                /mol_type="mRNA"
                /strain="FVB/N"
                /db_xref="taxon:10090"
                /clone="MGC:37241 IMAGE:4972485"
                /tissue_type="Kidney, normal. 5 month old male mouse."
                /clone_lib="NCI_CGAP_Kid14"
                /lab_host="DH10B"
                /note="Vector: pCMV-SPORT6"
gene            1..3454
                /gene="Pypaf5"
                /db_xref="LocusID:101613"
                /db_xref="MGI:2141990"

Qy    367  ATGAGAGGAGGGGCCGAGCGCGCCTACCGCTTCGTCAGGAGGAACCGAGACGCCGTTCGGCGC  426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12939  ATGAGAGGAGGGGCCGAGCGCGCCTACCGCTTCGTCGAAGGAGGAACCGAGACGCCGTTCGGCGC  12998

Qy    427  TGTGCTTCGTGCCCTTCGTG·GCTGGATCGTGTGCACCGCGTCTCGCGCCAGCTGGAGC  486
           ||||||||||||||||||||  ||||||||||||||||||||||||||||||||||||
Db  12999  TGTGCTTCGTGCCCTTCGTG·GCTGGATCGTGTGCACCGCGTCTCGCGCCCASCAGCTGGAC  13058

Qy    487  TCGGTGGGACCTGTCGCGACGTCCAAGACCACCACGTCAGTGTACCGCTTTTCATCA  546
           ||||||||||||||||||||||||||||||||||||||||  ||||||||||||||
Db  13059  TCGGTCGGGACCTGTCGCGGCACGTCCAAGACCACCACGTCAGTGTACCGCTT··CATCA  13118

Qy    547  CCAGGCGTTCTCAGTCGGCTTCGGCTCGGCTAGCCGACGGGCCCCCGGTTGCAGGGCGACCTGCGCA  606
           ||  ||||||||||||||||||||||||||||  |||||||||||||||||||||||||||
Db  13119  CCAGGCGTTCTCAGTCGGCTCGGCTCGGCTAGCCGGCGGGCCGCCCGGTTGCAGGGCGACCTGCGCA  13178

Qy    607  ATCTGTGCCGCCTGGCCCGGAGGCGTCCTCGGACGGGCGCAGTTGCCGAGAGG  666
           |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13179  ATCTGTGCCGCCTGGCCCGGAGGCGTCCTCGGACGGGCGCAGTTGCCGAGAGG  13238

Qy    667  AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGCACGCTGTTTCTCAGCAAAAAGG  726
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13239  AACTGGAGCAACTGGAGCTTGTGGCTCCAAAGTGCAGCACGCTGTTTCTCAGCAAAAAGG  13298

Qy    727  AGCTGCCGCGGGCGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACAGAGTTCCAGG  786
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13299  AGCTGCCGCGGGCGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACACCAGAGCTTCCAGG  13358

Qy    787  AGTTCCTCGCGGCACTGTCCCACCTGCTGAGGACGCGCCAGGACCCGGGCTG  846
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13359  AGTTCCTCGCGGCACTGTCCCACCTGCTGAGGACGCGCCAGGACCCGGGCTG  13418

Qy    847  GCGGCGTGGGGACACTCCTCGTGGGGACGCGCAGCCGCCAGCCGCCACTTGGTGCTCACCA  906
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13419  GCGGCGTGGGGACACTCCTCGTGGGGACGCGCAGCCGCCAGCCACCACTTGGTGCTCACCA  13478

Qy    907  CGGCTTCCTCTTCGGACTGCTGACGCGCGGATGCGGATGCGCGACATGCCGCCACTTCG  966
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13479  CGGCTTCCTCTTCGGACTGCTGACGCGCCGGATGCGGATGCGCGACATGCGGCCACTTCG  13538

Qy    967  GCTCATGGTTCAGAGGGTTCGAAGCAGAGGCCCTGCGGCGGTGGCTGCAGGGCAGGGAC  1026
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13539  GCTCATGGTTCAGAGGGTTCGAAGCAGAGGCCCTGCGGTGGCGGTGGCTGCAGGGACCAGGGAC  13598

Qy   1027  AGGGCTGCCGGAGTGGCACCAGAGGTGACCGGGGGCCAAAGGGTTCGAGGACACCG  1086
           ||||||||||||||||||||||||||||||||  ||||||||||||||||||||
Db  13599  AGGGCTGCCGCGGAGTGGCACCAGAGGTGACCGGGGGCCAAAGGGCTTCGAGGACACCG  13658

Qy   1087  AAGAGCCAGGAGCCGGAGGAGGAGGAGCCCAACTACCCCACTGGAGTTGCTGTACT  1146
           ||||||||||||||||||||||||||||||||||  |||||||||||||||||||
Db  13659  AAGAGCCAGGAGCCGGAGGAGGAGGAGCCCAACTACCCCACTGGAGTTGCTGTACT  13718

Qy   1147  GCCTGTACGAGACGACGGCGCAGGAGGACGCGTTTGCGCCCAAGGCCTGTGCCGGTTCCCGGAGC  1206
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13719  GCCTGTACGAGACGACGGCGCAGGAGGACGACGCGTTTGCGCCCAAGGCCTGTGCCGGTTCCCGGAGC  13778

Qy   1207  TGGCGCGCGCAGCCGGAGTGCGGTTCTGCCGCATGGACGTGTTCTGAGCTACTGCGTGA  1266
           ||||||||||||||||||||||||||||||||||||||||||||||  ||||||||||
Db  13779  TGGCGCGCGCAGCCGGAGTGCGGTTCTGCCGCATGGACGTGTGTCTGT·TGAGCTACTGCGTGA  13838

Qy   1267  GGTGCTGCCCTGCTGACAGCCACTGCGGTCACTGCGGGCTGCAGATAGTTGGTGCTGCCGGCAGG  1326
           ||||||||||||||||||||||||||||||||||||||||||||||  |||||||||||||
Db  13839  GGTGCTGCCCTGCTGACAGCGCACTGCGCGGTCACTGCGGATCAGATAGTTGGTGCTGCCGGCAGG  13898

Qy   1327  AGAAGAAGAAGAGCGCCTGACGAAAGGGGCCTTCCAGGCCAGCCAGCCCTGCGGGTGGCGGCAGTT  1383
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13899  AGAAGAAGAAGAGCGCCTCGCGGGGAAGCGCGGCCTCCAGGCCAGCCAGCCCTGCGGGTGGCGGCGGT  13955

RESULT 13
BC031139
LOCUS       BC031139                3454 bp    mRNA    linear    ROD 10-JUN-2003

CDS             418..1758
                /codon_start=1
                /product="Pypif5 protein"
                /protein_id="AAH31139.1"
                /db_xref="GI:21410402"
                /db_xref="LocusID:101613"
                /translation="MLKSAGTNGPRVQGLRTLCRLAREGILDHHKAQFSEEDLEKLK
                LRGSGVQTIFLNKKEIPGVLKTEVTYQFIDQSFQEFLAALSYLLEAERTPGTPAGGVQ
                KLLNSDAELRGHLALTTRPLFGLLNTEGLRDIGNHFGCVVPDHVKKDTLRWVQGQSHP
                KGPPVGAKKTAELEDIEDAEEEEEEDLNFGLELLYCLYETQEEDFVRQALSSLPEI
                VLERVRLTRMDLEVLNYCVQCCPDGQALRLVSCGLVAAKEKKKKKSLVKRLKGSQST
                KKQPPVSLLRPLCETMTTPKCH:SVLILSHCRLPDAVCRDLSEALKVAPALRELGLLQ
                SRLTNTGLRLLCE3LAWPKCQVKTLRMQ:PDLQEVINY:VIVLQQSPVLTTLDLSGCQ
                LPGVIVEPLCAALKGPKCSLKTLSLTSVELSENSLRDLQAVKTSKPDLSIIYSK"

BASE COUNT      884 a    909 c    902 g    759 t
ORIGIN

Query Match        35.1%;  Score 944.8;  DB 10;  Length 3454;
Best Local Similarity   69.4%;  Pred. No. 4.3e-138;
Matches 1426;  Conservative   0;  Mismatches 597;  Indels  33;  Gaps   0;

```
Db  1861  ||  ||  |||  ||| ||| |||  ||  ||
          ACCTCCCAT-CAAAAGCAAGAAGATGCCAGCTTCTCCCTCCAGGCAAGTCTTTCAAGCC  1920
Qy  2069  GAGGCCCAGACAGGCATGTCGGAGGCCCAGACACGGCACCCTGCCCCGTCCAGGACAGG  2127
          |- |||  ||| ||                    ||| ||  ||
Db  1921  ACAGAAG--GGCAAGCAAAAGACCCAGTAGATACAGGCCCTTGATAGCTCCTGGATAGA  1978
Qy  2128  CCCAGGACCTGCCCCTCTTCCACACCTGGGTACCCTTCTCCCCAGCCCACCACTA  2187
          |||
Db  1979  CCTGGCAGTGCCTGCCCCTGCACTCTGCGGTTGGTCTTCCTCAGTCCCACACCAG  2038
Qy  2188  CTCCACCACCTTCCT  2203
          |||  ||  |||
Db  2039  CAATGCCTTCCTTTCT  2054
```

```
BASE COUNT      588 a    614 c    642 g    510 t
ORIGIN

Query Match:                33.9%;     Score 911.2; DB 10;    Length 2354;
Best Local Similarity 69.3%;     Pred. No. 8.3e-133;
Matches 1381;    Conservative   0; Mismatches 578;   Indels   33;   Gaps    9;
```

```
Qy  221   GCTGCTGCCCACGGCGCCCTCCTGCTGGTGACCACCGGCACCCGCGGCCCCCGGGAGGCTGCA   280
          |||||||||||| ||  ||  || |||||||
Db  2     GCTGTTACCAGGAGCTCGC-TGCTAGTGACTACACCACCATGCCCGTCACAGGAGGCTGCA   61

Qy  281   GGGCCGCCGTGTTCCCCGGAGCGTGCGCCAGAGTGGCCGAGGGCGGGCCGAGCCGCCTTCGT   340
          || ||||  ||| || |||       ||| |||| || || ||
Db  62    GGGCAGATTGTGCTCGCCACAGTGCGCAGAAATACGCGGCTCTTCAGACAGAACAAAAA   121

Qy  341   GAAGTATTCTACAAGTTCTTCCGGGATGAGAGGAGGGCCGAGCCGCCTTACCGCTTCGT   400
          ||||||||||||||||||||||
Db  122   AAAGTATTCTCAAGTTCTTCCGGGACGAGAGGAGGGCAGAGCCGCCTTACCGCTTCGT   181

Qy  401   GAAGGAACGAGAGCGTGTTCGGCTGTTGCTTCGTGTGCCCTTCGTGTGTGGGATCGTGTG   460
          |||||||||||||||| |||  ||||||||||||
Db  182   GAAAGAGAGAACGAGAGCGTGTTGCACTGTGCTTTGTGCCCTTCGTGTGCTGGATCGTGTG   241

Qy  461   CACCGTGCTGCCCAGCAGCTGGAGCTCGGTCGCGCAGTGCCAAGACCACC   520
          |||||||  ||| |||||||||||||||||||||||
Db  242   CACAGTGCTGCAGCAGCTAGAGCTCCTGCTCTCCAAGACCACC   301

Qy  521   CACGTCAGTGTACCTGCTTTTCATCACCACGCTTCTGAGCTCGGCTCCGTAGCCGACGG   580
          |||||||||||||||||||| |||||
Db  302   TACTTCTGTGTACCTGCTCTTCATCACCAGCATGCTGA---AGTCTGCAGGCACCAATGG   358

Qy  581   GCCCCGGTTGCAGGGCCGACCTGCCAATCTGTGCCCGGGCCCTGGCCTGCCGCGAGGGCGTCCTCG-   639
          |||| ||       ||||||||||||||||||||||
Db  359   ACCCCGGGTTCAGGGAGAGCTGAGACGCTGTGTGCCGGGGCATCCTGGA   418

Qy  640   --GACGCAGGGGCCGCAGTTTGCCGAGAAGGAACTGGAGAACTGGAGCTTCGTGGCTCCAA   697
          || |||||||||||||||||||||||||||||||||||
Db  419   TCATCATAAAGACACAATTCTCAGAAGAGGAACCTGGAGAAATTGAAGCTTCGGGGGTTCCCA   478

Qy  698   AGTGCAGCAGCGCTGTTTCTCAGCAAAAAGGAGCTGCGGGGTGCTGGAGACAGAGGTCAC   757
          ||||||||||||||||||||||||
Db  479   AGTTCAGACAATATTTCTCAACAACAAGAGATACCAGGAGTGCTAAAAAC=GAGGTCAC   538

Qy  758   CTACCAGTTCATCGACCAGAGCTTCCAGGAGTTCCTGGCGGCACTGTCCTACCTGCTGGA   817
          ||||||||||||||||||||||||||| ||||||||||
Db  539   CTACCAGTTCATTGACCAGAGCTTCCAGGAGTTCCTGGCTGCATTGTCATACCTACTAGA   598

Qy  818   GGACGGCGGGGTGGCCTGCCCAGGACCGCGGCGCCCGGGGGCGGTTGGGACACTCCTGCGTGGGAGGGC   877
          |||| ||      || ||||
Db  599   AGCTGAGCGAACTCCGGGACGGCCGGCCGCAGCGGTGCAGAAGCTCCTGAACTCTGACGGC   658

Qy  878   CCAGCGCCACAGCCACTTGGTGCTCACACGGCGCTTCACCGGGACTCTGAGCGCGGGA   937
          |||||||||||||||| |||||
Db  659   GGAGCTACGTGGTCATCTTGCACTGACCACCGATTCCTTTTGGCTGCTAAATACAGA   718

Qy  938   GCGGATGCGGCGACATCGAGCGGTTCCACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGA   997
          |||||||||||      |||||||||||||||||
Db  719   GGGGCTTCGTGACATTGGAAACCATTTTGCTGTTCGGTGTGGTGCCAGATCATGTGAAAAAGGGA   778

Qy  998   GGCCCTGCGTGGGTGCGCAGGGACAGGGACGTGCAGGGTGTGCACCAGGTGAC   1057
          |||||||||||||| ||||||||||
Db  779   CACCCTGCCGTGGGT------ACAAGGACAGACCAGCATCCGAGGACGACCCCAGGGGGGC   832

Qy  1058  CGAGGGGGCCAAAGGGCTGGAGCAGCCCAGAGAGGCCAGGAGGAGGGAGAGGAGAGGA   1117
          |||||||||
Db  833   AAAAAAGACTGCTGAGCTGGAGGACATCGAGGACGACGCCAGGAGGAGGAGGAGGAGGAGGA   892
```

Db    1959 GCCTCCTTTCT 1969
              || ||||| .||
           1959 GCCTCCTTTCT 1969

RESULT 15
AP006286
LOCUS       AP006286              175416 bp    DNA     linear   HTG 26-MAR-2003
DEFINITION  Homo sapiens chromosome 11 clone RP13-25N22 map 11p, WORKING DRAFT
            SEQUENCE, 22 unordered pieces.
ACCESSION   AP006286
VERSION     AP006286.1  GI:29243346
KEYWORDS    HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Homo sapiens genomic DNA of 11p
  JOURNAL   Published Only in Database (2003)
REFERENCE   2  (bases 1 to 175416)
  AUTHORS   Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
            Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
  TITLE     Direct Submission
  JOURNAL   Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical
            and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
            1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
            Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT     ---------- Genome Center
            Center: RIKEN Genomic Sciences Center(GSC)
            Center code: RIKEN
            Web site: http://hgp.gsc.riken.go.jp/
            Contact: hattori@gsc.riken.go.jp
            ---------- Project Information
            Center project name: HumDraft11
            Center clone name: RP13-25N22
            ---------- Summary Statistics
            Sequencing vector: PCR products; 100% of reads
            Chemistry: Dye-terminator ET-amersham; 100% of reads
            Assembly program: Phrap; version 0.990329
            Consensus quality: 160074 bases at least Q40
            Consensus quality: 167722 bases at least Q30
            Consensus quality: 171662 bases at least Q20
            Insert size: 173316; sum-of-contigs
            Quality coverage: 7.92x in Q20 bases; sum-of-contigs
            ----------
            NOTE: This is a 'working draft' sequence. It currently consists of
            22 contigs. The true order of the pieces is not known and their
            order in this sequence record is arbitrary. Gaps between the
            contigs are represented as runs N, but the exact sizes of the gaps
            are unknown. This record will be updated with the finished sequence
            as soon as it is available and the accession number will be
            preserved
                     1      33131 contig of   33131 bp in length
                 33232      54098 contig of   20867 bp in length
                 54199      72037 contig of   17839 bp in length
                 72138      83277 contig of   11140 bp in length
                 83378      98599 contig of   15222 bp in length
                 98700     105513 contig of    6814 bp in length
                105614     116229 contig of   10616 bp in length
                116330     121964 contig of    5635 bp in length
                122065     127627 contig of    5563 bp in length
                127728     131971 contig of    4244 bp in length
                132072     137506 contig of    5435 bp in length
                137607     143829 contig of    6223 bp in length
                143930     148539 contig of    4610 bp in length
                148640     153148 contig of    4509 bp in length
                153249     157509 contig of    4261 bp in length
                157610     160561 contig of    2952 bp in length
                160662     164778 contig of    4117 bp in length
                164879     168269 contig of    3391 bp in length

Qy 1118 G--CCCAACTACCCACTGGAGTTGCTGTACTGCC-GTACGAGAGACGCCAGGAGGACGCGGT 1174
             ||||||  |||||||  |||||||||| ||  | |||||||||||||||||||||||
Db  893 GGACCTCAATTTGGACTGGAGCTGCTGCTGCACTGCCTGCTGTATGGAGCACCAGGAGGAGGGATT: 952

Qy 1175 TGTGCGGCAAGCCCTGTCCGGTTCCCGGACTGGCGCTGCGGCTTCTGCCG 1234
             ||||||| ||| |||||||  ||||||||| |||||| |||||||||
Db  953 TGTTCGGACAGGGTCCTCAGCACGCCCTTCTCCAGAGATAGTACTGGAGCGAGAGTCGCCAGGCTGACCCG 1012

Qy 1235 CATGGACGTGGCTGTTCTGAGCTACTGCGCGAGGTGCTGCCCTGCTGGCAGGCACTGCG 1294
             ||||||  |||||| ||||||  |||| ||||||||| || |||||||  |||||||||
Db 1013 CATGGACTTCGAATTACTGCGTGCAGTGCTGCGTGCCGCAGATGGCCAGGGCTCTGAG 1072

Qy 1295 GCTGATCAGCTGCAGATTGGTTGGCTGGCTGGCGGCAGGAGGAAGAAGAGCCTGGGGAAGCG ------AGAAGAG 1354
             ||||| |||||||||||||||||||| |||||||||| |||||||||||||||||||||||
Db 1073 ACTGGTGAGCTGTGGGGCTGGTGGTGGCGCAACTGGGCAAGGAGGAAGAAGAAGAGA------AGAAGAG 1126

Qy 1355 GCTCCAGGCCAGCCTGGGTGGCTGGGCGGCAGTTCTCAAGGCACCAAAAACAACTGCCAGCCTC 1414
             || |||    ||||||||   || ||||||  |||||||||||||||||||||||||||||||
Db 1127 CCTGGTGAAGCCGGCTGAAGGGC------TCTCAAAGCACCAAGAAACAACCCCAGTCTC 1180

Qy 1415 CCTTCTCATCCACTCTTTCAGGCAATGACTTGCCCACTGTGCCATCTGAGCAGCCTCAC 1474
             |||||| |||||||  ||||||| |||||  |||||||||||  |||| |||||||||||
Db 1181 CTTGCTCGTCGTCTGTGAGACAATGACTACCCCCAAATGTCATCTCTGAGTGTTCTGAT 1240

Qy 1475 GCTGTCCACTGCAAACTCCTGACGCGGGTCTGCGGAGACCTTTCTGAGGCCCTGAGGGC 1534
             | |||||||||||||| ||||||||| |||||||| |||||||||||||||||||||
Db 1241 CTTGTCACACTGCAGACTCCCTGATGCAGTTTGCCCGAGACCCTTTCCGAGGCCCTGAAGGT 1300

Qy 1535 AGCCCCCGCACTGACGGAGCCTGGGCTTCCTCCACAACAGGCTCAGTGAGGCGGGACTGCG 1594
             ||| |||| | || ||||||||||||| ||| ||||  ||||||| |||||||||||||
Db 1301 AGTCCGCCCAAGGGAGTTGGGCCTCCTAAGGACCCTCACCAACACAGGCTCGCG 1360

Qy 1595 TATGCTGAGTGAGGGCCTAGCCTGGCCGCGCAGTGCCAGGTGCAGACGGTCAGGGTACAGCT 1654
             | |||||||  || |||  |||||||||||| ||| |||||||| |||||||||||| ||||
Db 1361 TTTACTGTGTGCGAAGGCCTGGGCTTGGCCCAAGCTGGGCCCAAGTGAAGACACTCAGGATGCAGCT 1420

Qy 1655 GCCTGACCCCCAGCGAGGGCTCCAGTACCTGG-ATGCTCGGCAGAGCCCCCCT 1714
             ||| ||  |||||| ||| ||||||| | ||| |  ||||| |||||||||
Db 1421 GCCTGAACTCCAAGGAGTGATCAATTTACTTGGTCATCGGCGTCCAGCGAGCCCAGTCCT 1480

Qy 1715 GACCACCCTGGATCTCAGCGCGGTGCCAAACTGCCCCGCCCCCATGGTGACCTACCTGTGTGC 1774
             || |||||||  ||||||| |||||| |||||||||| || |||||||||||||||||||
Db 1481 GACCACTTGGACCTCAGGTGCGTGTCAGTGCGTCATGTGGAACCTCTGGTGCCCTGGTGAGGCTGAG 1540

Qy 1775 AGTCTGCAGCACCAGGGATGCGGGCCTGCAGACCCTCAGTCGGCCTCTGTGGAGGCTGAG 1834
             |||||||||| |||||  ||| ||||||| || |||||||||| ||||||||||||| |||
Db 1541 AGCCCTGAAGCACCCTAAATGCTCCCTAAAGACCCTCAGTCGACCTCTGTGTGGAGTGAG 1600

Qy 1835 CGAGCAGTCACTACAGGAGCTTCAGGCTGTGAAGAGCAGCAAAGCCGGATCTGGTCATCAC 1894
             || |  ||  || || ||   || ||||||||||||||| ||||| || ||| || ||||
Db 1601 TGGAACTCACTGAGGGATTTCAAGCTGTGAAGACTCAAAGCCAGATCTGTCCATCAT 1660

Qy 1895 ACACCCAGCGCGTGGAGCGCGCCACCCAACCTCCAGGAACTCATCTGACCTTCTGAGG 1954
             |||||||||||  |||||| ||||||||| |||||||||||||||||| ||||||||||
Db 1661 ATATTCAAAATAAGGCACACCTCACCGGAACTCCGGAAGG-GTGAGGCAGTGTCTTCTGAAG 1719

Qy 1955 CTCTGGTGGCCAGAGCAGGCATGTGGGAGGCGGTGGAAGACCT---AGTCAAAGTCCCCGTGGAGAGAACGG 2011
             ||| ||  |||| |||| ||  |||| |||||||   ||||||   ||||||||||||| ||||||||||||
Db 1720 CTCCGGAGCCCAGAACACGATGGAAGATGCTGTAGCCCAAGTGCTGTAGCCCAAGTCCCTCTGAATATCGACCT 1779

Qy 2012 CCCATTCCAAGGGCAGGAGGAGATATTGCTTCTCCGGCCTTTGAGCCCGAGAGG 2071
             |||| |||||||||||||  |||||  ||| |  |||||||||||||||||||||
Db 1780 CCCATTCAAAGGGCAGGAGGATGCCAGTGGTTCCCTCCAGGCAAGTCTTTCAGGCACAG 1839

Qy 2072 CCGCAGACAGGCATGTGGGAGGCCGCACCGCACCCTGCCCCGTCCAGGACAGGCCCA 2131
             || ||||||| || ||||||| |||| |||||  |||||| ||||||||||| ||||
Db 1840 AAG--GGCAAGCAAAGACCCAGCTAGATACAGGGCCCTGATAGTCCTGGATAGAGACCTG 1897

Qy 2132 GGACCTGCCCCTCTCCAACCTGGGGTACCCCTTCTCCCCCAGCCCCACCCACTACTCC 2191
             ||||||||||| ||| |  ||||||| |||||||| |||||| |||||||||||||||
Db 1898 GCACGTGCCCTGCCCTGACACTCTGCGGTTGCTCTGCGCCCTCCTCAGTCCCACAACCAGCAAT 1957

Qy 2192 ACCCACTTCCT 2203
           2192 ACCCACTTCCT 2203

```
        168370  170689 contig of   2320 bp in length
        170790  172833 contig of   2044 bp in length
        172934  174155 contig of   1222 bp in length
        174256  175416 contig of   1161 bp in length
*  NOTE: This is a 'working draft' sequence. It currently
*  consists of 22 contigs. The true order of the pieces
*  is not known and their order in this sequence record is
*  arbitrary. Gaps between the contigs are represented as
*  runs of N, but the exact sizes of the gaps are unknown.
*  This record will be updated with the finished sequence
*  as soon as it is available and the accession number will
*  be preserved.
             1     33131: contig of  33131 bp in length
        33132     33231: gap of    100 bp
        33232     54098: contig of  20867 bp in length
        54099     54198: gap of    100 bp
        54199     72037: contig of  17839 bp in length
        72038     72137: gap of    100 bp
        72138     83277: contig of   1140 bp in length
        83278     83377: gap of    100 bp
        83378     98599: contig of  15222 bp in length
        98600     98699: gap of    100 bp
        98700     105513: contig of   6814 bp in length
       105514    105613: gap of    100 bp
       105614    116229: contig of  10616 bp in length
       116230    116329: gap of    100 bp
       116330    121964: contig of   5635 bp in length
       121965    122064: gap of    100 bp
       122065    127627: contig of   5563 bp in length
       127628    127727: gap of    100 bp
       127728    131971: contig of   4244 bp in length
       131972    132071: gap of    100 bp
       132072    137506: contig of   5435 bp in length
       137507    137606: gap of    100 bp
       137607    143829: contig of   6223 bp in length
       143830    143929: gap of    100 bp
       143930    148539: contig of   4610 bp in length
       148540    148639: gap of    100 bp
       148640    153148: contig of   4509 bp in length
       153149    153248: gap of    100 bp
       153249    157509: contig of   4261 bp in length
       157510    157609: gap of    100 bp
       157610    160561: contig of   2952 bp in length
       160562    160661: gap of    100 bp
       160662    164778: contig of   4117 bp in length
       164779    164878: gap of    100 bp
       164879    168269: contig of   3391 bp in length
       168270    168369: gap of    100 bp
       168370    170689: contig of   2320 bp in length
       170690    170789: gap of    100 bp
       170790    172833: contig of   2044 bp in length
       172834    172933: gap of    100 bp
       172934    174155: contig of   1222 bp in length
       174156    174255: gap of    100 bp
       174256    175416: contig of   1161 bp in length.
FEATURES             Location/Qualifiers
     source          1..175416
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p"
                     /clone="RP13-25N22"
     misc_feature    1..33131
                     /note="assembly_fragment clone_end:SP6 vector_side:left"
     misc_feature    33232..54098
                     /note="assembly_fragment"
     misc_feature    54199..72037
                     /note="assembly_fragment"
     misc_feature    72138..83277
                     /note="assembly_fragment"
     misc_feature    83378..98599
                     /note="assembly_fragment"
```

```
     misc_feature    98700..105513
                     /note="assembly_fragment clone_end:T7 vector_side:left"
     misc_feature    105614..116229
                     /note="assembly_fragment"
     misc_feature    116330..121964
                     /note="assembly_fragment"
     misc_feature    122065..127627
                     /note="assembly_fragment"
     misc_feature    127728..131971
                     /note="assembly_fragment"
     misc_feature    132072..137506
                     /note="assembly_fragment"
     misc_feature    137607..143829
                     /note="assembly_fragment"
     misc_feature    143930..148539
                     /note="assembly_fragment"
     misc_feature    148640..153148
                     /note="assembly_fragment"
     misc_feature    153249..157509
                     /note="assembly_fragment"
     misc_feature    157610..160561
                     /note="assembly_fragment"
     misc_feature    160662..164778
                     /note="assembly_fragment"
     misc_feature    164879..168269
                     /note="assembly_fragment"
     misc_feature    168370..170689
                     /note="assembly_fragment"
     misc_feature    170790..172833
                     /note="assembly_fragment"
     misc_feature    172934..174155
                     /note="assembly_fragment"
     misc_feature    174256..175416
                     /note="assembly_fragment"
BASE COUNT    37167 a  49392 c  49417 g  37339 t   2101 others
ORIGIN

Query Match          31.2%;  Score 838.4;  DB 2;  Length 175416;
Best Local Similarity 99.9%;  Pred. No. 9.9e-122;
Matches  839;  Conservative   0;  Mismatches   0;  Indels   0;  Gaps   0;

Qy       18:1 CAGTCTGGCCTCTGTGGAGCCAGCAGTCACTACAGGAGCTTCAGGCTGTGAAGAG 1870
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101864 CAGTCTGGCCTCTGTGGAGCCAGCAGTCACTACAGGAGCTTCAGGCTGTGAAGAG 101923

Qy       1871 AGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGCCACCCACAACCTCCCAA 1930
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101924 AGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGCCACCCACAACCTCCCAA 101983

Qy       1991 GGAACTCATCTCGACCTTCTGAGGCTCTGGTGGCCAAGAGCAGGGTGGAAGACCCTAGTCA 1990
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     101984 GGAACTCATCTCGACCTTCTGAGGCTCTGGTGGCCAAGAGCAGGGTGGAAGACCCTAGTCA 102043

Qy       1991 AAGTCCCGTGTGGAGAGAACGGCCCATTCCAAGGGCAGGAGCATGTGGGAGGCCCAGACACGGCACCCT 2050
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102044 AAGTCCCGTGTGGAGAGAACGGCCCATTCCAAGGGCAGGAGCATGTGGGAGGCCCAGACACGGCACCCTTTG 102103

Qy       211? GCCCCGTCCAGGACAGGCCCAGGACCTGCCCCTCTCTCCCACACCTGGGGTACCCCTTTCTC 2170
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102164 GCCCCGTCCAGGACAGGCCCAGGACCTGCCCCTCTCTCCCACACCTGGGGTACCCCTTTCTC 102223

Qy       2171 CCCCAGCCCACCACTACTCCACCTCCACCC 2230
              ||||||||||||||||||||||||||||||
Db     102224 CCCCAGCCCACCACTACTCCACCTCCACCCTCCTGAGACCCTCCAGCCATTCCCCT 102283

Qy       2231 TGAAAACACCCCCGACCCCAAGCCACAATAATGACAGGAGAGCTCCAATTAACTAAGC 2290
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     102284 TGAAAACACCCCCGACCCCAAGCCACAATAATGACAGGAGAGCTCCAATTAACTAAGC 102343
```

```
Qy      2291 ACCTACCTGGCGGCAGAGATAACCCTTCACTGCCTGATCCCCATCTGCAGTGTGGCCCAAC 2350
             ||||||||||| ||||||| |||||||||||||||||||||||||||||||||||||||||
Db    102344 ACCTACCTGGGGGCAGAGAATAACCCTTCACTGCCTGATCCCCATCTGCAGTGTGGCCCAAC 102403

Qy      2351 AGCCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGTCCCTCCCTGTTA 2410
             |||||||||||||| |||||||||||||||||||||||||||||||||| ||||| || |
Db    102404 AGCCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGTCCCTCCCTG-TA 102462

Qy      2411 GGAATGAGACCATCCCTGAGGCTATGGCCCAGGCCCACAGGCGTCCAGTGTCTGAGATCT 2470
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102464 GGAATGAGACCATCCCTGAGGCTATGGCCCAGGCCCACAGGCGTCCAGTGTCTGAGATCT 102523

Qy      2471 TTGGGAAGGGAGACTAGGGCAGTGGAGACAGCGCAGAACCCCGTGCTGGGTGGGAAGC 2530
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102524 TTGGGAAGGGAGACTAGGGCAGTGGAGACAGCGCAGAACCCCGTGCTGGGTGGGAAGC 102583

Qy      2531 ATGACCACATGGTGGGTGAGCAGCCCCCATGCACTGACGGTAAATTCCCCTGTGGACTCA 2590
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    102584 ATGACCACATGGTGGGTGAGCAGCCCCCATGCACTGACGGTAAATTCCCCTGTGGACTCA 102643

Qy      2591 TTTCTGTTGGTTTCTATTACACCTGGCCAGGCGTGGTACAATACAGGTCGGTGCTCACAA 2650
             ||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||
Db    102644 TTTCTGTTGGTTTCTATTACACCTGGCCAGGCG-GGTACAATACAGGTCGGTGCTCACAA 102703
```

Search completed: October 3, 2003, 11:59:29
Job time : 9833 secs

```
               GenCore version 5.1.6
               Copyright (c) 1993 - 2003 Compugen Ltd.

GM protein - protein search, using sw model

Run on:       October 2, 2003, 17:35:13 ; Search time 86 Seconds
                                           (without alignments)
                                         1151.536 Million cell updates/sec

Title:        US-10-029-347-2
Perfect score: 3218
Sequence:     1 MLAQPQRLLFILDGADELPA..........ITHPALDGHPQPPKELISTF 625

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:     1:07863 seqs, 158726573 residues

Total number of hits satisfying chosen parameters:      1107863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :    A_Geneseq 19Jun03:*
           1: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1980.DAT:*
           2: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1981.DAT:*
           3: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1982.DAT:*
           4: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1983.DAT:*
           5: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1984.DAT:*
           6: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1985.DAT:*
           7: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1986.DAT:*
           8: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1987.DAT:*
           9: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1988.DAT:*
          10: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1989.DAT:*
          11: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1990.DAT:*
          12: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1991.DAT:*
          13: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1992.DAT:*
          14: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1993.DAT:*
          15: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1994.DAT:*
          16: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1995.DAT:*
          17: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1996.DAT:*
          18: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1997.DAT:*
          19: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1998.DAT:*
          20: /SIDS1/gcgdata/geneseq/geneseq-embl/AA1999.DAT:*
          21: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2000.DAT:*
          22: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2001.DAT:*
          23: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2002.DAT:*
          24: /SIDS1/gcgdata/geneseq/geneseq-embl/AA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Match | Query Length | DB | Length | ID | Description |
|---|---|---|---|---|---|---|---|
| 1 | 3218 | 100.0 | 625 | 23 | ABG78454 | Human leucine-rich |
| 2 | 3218 | 100.0 | 891 | 22 | AAE04546 | Human G-protein co |
| 3 | 3144 | 97.7 | 676 | 23 | ABU65075 | Human NOV18a prote |
| 4 | 2882 | 89.6 | 1551 | 22 | AAU00023 | Human activated 7- |
| 5 | 2882 | 89.6 | 1851 | 23 | AAO17860 | Pyrin domain conta |
| 6 | 2239 | 69.6 | 597 | 22 | ABB11735 | human vasopressin |
| 7 | 1578 | 49.0 | 481 | 14 | AAR33389 | AII/AVPv2 receptor |
| 8 | 1120 | 34.8 | 321 | 23 | ABB9C173 | Human polypeptide |
| 9 | 1056.5 | 32.8 | 329 | 22 | AAM89960 | Human immune/haema |

| No. | Score | % | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 10 | 825.5 | 25.7 | 1061 | 23 | AAO15590 | Human PYRIN-8 prot |
| 11 | 825 | 25.6 | 1099 | 23 | AAO17857 | Pyrin domain conta |
| 12 | 800.5 | 24.9 | 1034 | 22 | AAE07514 | Human PYRIN-1 prot |
| 13 | 800.5 | 24.9 | 1034 | 24 | ABU08503 | Huma PYRIN-1 prote |
| 14 | 762 | 23.7 | 1429 | 22 | AAB62571 | Human CARD-7 polyp |
| 15 | 762 | 23.7 | 1429 | 23 | ABG97969 | Human leucine-rich |
| 16 | 762 | 23.7 | 1429 | 23 | ABG78472 | leucine-rich repea |
| 17 | 762 | 23.7 | 1429 | 23 | ABB77916 | Human leucine-rich |
| 18 | 762 | 23.7 | 1429 | 23 | ABG71631 | Human caspase recr |
| 19 | 762 | 23.7 | 1429 | 24 | ABG71633 | Human caspase recr |
| 20 | 762 | 23.7 | 1454 | 22 | AAY72673 | Human NAC beta iso |
| 21 | 762 | 23.7 | 1473 | 22 | AAE06758 | Human G-protein co |
| 22 | 762 | 23.7 | 1473 | 22 | AAY72669 | Human NB-ARC and C |
| 23 | 762 | 23.7 | 1473 | 22 | AAY7271 | Human NAC beta iso |
| 24 | 760 | 23.6 | 1429 | 23 | ABG78455 | Human caspase recr |
| 25 | 760 | 23.6 | 1429 | 23 | AAO17855 | Pyrin domain conta |
| 26 | 716 | 22.2 | 1424 | 22 | AAY72674 | Human NAC gamma or |
| 27 | 712.5 | 22.1 | 1397 | 23 | AAY72670 | Human NB-ARC and C |
| 28 | 712.5 | 22.1 | 1442 | 22 | AAY72671 | Human NB-ARC and C |
| 29 | 691 | 21.5 | 920 | 23 | ABP53254 | Human MDDT-13 prot |
| 30 | 660 | 20.5 | 952 | 22 | ABG03924 | Novel human diagno |
| 31 | 650 | 20.2 | 980 | 23 | ABB98244 | Human CGD SEQ ID |
| 32 | 650 | 20.2 | 980 | 23 | AAM50329 | Human nucleotide b |
| 33 | 641 | 19.9 | 449 | 23 | ABB77910 | Amino acid sequenc |
| 34 | 635 | 19.7 | 1022 | 23 | AAO17862 | Pyrin domain conta |
| 35 | 634 | 19.7 | 1049 | 22 | AAG65894 | Amino acid sequenc |
| 36 | 619.5 | 19.3 | 1033 | 22 | AAE07513 | Human nucleotide b |
| 37 | 619.5 | 19.3 | 1033 | 23 | ABG78456 | Human nucleotide b |
| 38 | 619.5 | 19.3 | 1033 | 24 | ABU08502 | Huma pyrin domain: |
| 39 | 616.5 | 19.2 | 1062 | 22 | AAG68895 | Amino acid sequenc |
| 40 | 616.5 | 19.2 | 1062 | 23 | AAO17856 | Pyrin domain conta |
| 41 | 598.5 | 18.6 | 858 | 23 | AAM50328 | Human nucleotide b |
| 42 | 598.5 | 18.6 | 919 | 23 | ABP43483 | Human secreted pro |
| 43 | 598.5 | 18.6 | 919 | 24 | ABJ19919 | NOVX related prote |
| 44 | 598.5 | 18.6 | 983 | 23 | AAO17870 | Pyrin domain conta |
| 45 | 598.5 | 18.6 | 994 | 23 | AAO15593 | Human PYRIN-3 prot |

ALIGNMENTS

```
RESULT 1
ABG78454
ID   ABG78454 standard; Protein; 625 AA.
XX
AC   ABG78454;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Human leucine-rich repeat small intestine I (HLRRSI1).
XX
KW   Human; human leucine-rich repeat, small intestine I; HLRRSI1; asthma;
KW   proliferative disorder; gastrointestinal disorder; renal disorder;
KW   neural disorder; reproductive disorder; calcium regulation; apoptosis;
KW   immune system; anaemia; human immune deficiency virus; HIV; cancer;
KW   blood coagulation disorder; autoimmune disorder; allergic reaction;
KW   inflammatory condition; cardiovascular disorder; ischaemia;
KW   neurological disorder; infectious disease; cytokine production;
KW   expressed sequence tag; EST.
XX
CS   Homo sapiens.
XX
PN   WO200261086-A2.
XX
PD   08-AUG-2002.
XX
PF   23-DEC-2001; 2001WO-US049739.
XX
PR   22-DEC-2000; 2000US-257774P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Feder J, Ramanathan C, Mintier G;
```

```
Qy       481  LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGM 540
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       481  LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGM 540

Qy       541  LRQSPALTLDLSGCQLPAPMVTYLCAVLQHQCGLQTLSLASVELSEQSLQELQAVKRA 600
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       541  LRQSPALTLDLSGCQLPAPMVTYLCAVLQHQCGLQTLSLASVELSEQSLQELQAVKRA 600

Qy       601  KPDLVITHPALDGHPQPPKELISTF 625
              |||||||||||||||||||||||||
Db       601  KPDLVITHPALDGHPQPPKELISTF 625
```

RESULT 2

AAE04546

```
ID   AAE04546 standard; Protein; 891 AA.
XX
AC   AAE04546;
XX
DT   04-SEP-2001 (first entry)
XX
DE   Human G-protein coupled receptor-2 (GCREC-2) protein.
XX
KW   Human; G-protein coupled receptor-2; GCREC-2; gene therapy; cirrhosis;
KW   transgenic animal; proliferative disorder; actinic keratosis; hepatitis
KW   nephrotropic; cancer; breast; bladder; bone marrow; brain; uterus;
KW   leukaemia; adenocarcinoma; lymphoma; melanoma; myeloma; epilepsy; stroke;
KW   neurological disorder; neurological disorders, e.g. Parkinson's disease; nausea;
KW   Huntington's disease; multiple sclerosis; dementia; angina pectoris;
KW   central nervous system disorder; cardiovascular disorder; hypertension;
KW   atherosclerosis; congestive heart failure; gastrointestinal disorder;
KW   dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;
KW   pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
KW   inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
KW   Addison's diseases; allergy; asthma; diabetes mellitus; antithyroid;
KW   atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
KW   rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
KW   metabolic disorder; obesity; nootropic; protozoacide; virucide.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Binding-site    202..209
FT                   /label= P loop
FT                   /note= "ATP/GTP binding site"
XX
PN   WO20014228B-A2.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US33382.
XX
PR   10-DEC-1999; 99US-0172852.
PR   22-DEC-1999; 99US-0171732.
PR   14-JAN-2000; 2000US-0176148.
PR   21-JAN-2000; 2000US-0177331.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Burford N,  Baughn MR,  Au-Young J,  Yang J,  Lu DAM,  Reddy R;
XX
DR   WPI; 2001-381635/40.
DR   N-PSDB; AAD08836.
XX
PT   New human G-protein coupled receptor polypeptides for diagnosing,
PT   preventing, and treating cell proliferative, neurological,
PT   cardiovascular, gastrointestinal, autoimmune and metabolic disorders -
XX
PS   Claim 1; Page 129-131; 175pp; English.
XX
CC   The present sequence is human G-protein coupled receptor-2 (GCREC-2)
CC   protein. GCREC is useful in somatic or germline gene therapy to correct
CC   a genetic deficiency, to express a conditionally lethal gene product and
```

```
XX
DR   WPI; 2002-619252/66.
DR   N-PSDB; ABS63485.
XX
PT   New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT   their fragments and homologues, useful for preventing, treating and
PT   ameliorating medical conditions, e.g. proliferative, gastrointestinal,
PT   or renal disorders -
XX
PS   Claim 20; Figure 2; 336pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules (I) encoding
CC   human leucine-rich repeat small intestine I (HLRRSI1) polypeptides.
CC   The nucleic acid molecules and polypeptides are useful for preventing,
CC   treating and ameliorating medical conditions, such as proliferative,
CC   gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC   related to aberrant calcium regulation or apoptosis modulation, either
CC   directly or indirectly. They are also useful for treating, preventing
CC   and/or diagnosing diseases, disorders and/or conditions of: immune system
CC   by activating or inhibiting the proliferation, differentiation, or
CC   mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC   anaemia; immunologic deficiency syndromes, e.g. human immune deficiency
CC   virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC   e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC   myasthenia gravis; asthma or allergic reactions; inflammatory
CC   conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC   e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC   ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC   Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC   pneumonia, or viral, bacterial, and fungal infections. The HLRRSI-
CC   polypeptides are useful for modulating cytokine production, antigen
CC   presentation, or other processes such as boosting immune responses.
CC   ABG78454-ABG78474 represent HLRRSI1 amino acid sequences and related
CC   amino acid sequences of the invention.
XX
SQ   Sequence  625 AA;
```

```
Query Match             100.0%;  Score 3219;  DB 23;  Length 625;
Best Local Similarity   100.0%;  Pred. No. 5.3e-314;
Matches  625;  Conservative    0;  Mismatches    0;  Indels    0;  Gaps    0;

Qy         1  MLAQPQRLFLILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 60
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         1  MLAQPQRLFLILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 60

Qy        61  RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        61  RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120

Qy       121  VPFVCWIVCTVLRQQELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLCGDLRNLC 180
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       121  VPFVCWIVCTVLRQQELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLCGDLRNLC 180

Qy       181  RLAREGVLGRRACFAEKELECLELRGSKVQTLFJSKKELPGVLETEVTYQFIDQSFQETL 240
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       181  RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFJSKKELPGVLETEVTYQFJDQSFQEFL 240

Qy       241  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 300
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       241  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 300

Qy       301  VSERVKQEALRWVQGGQGQGPGVAPEVTEGAKGLEDTEPPEEEEGEEPNYPLELLYCLY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       301  VSERVKQEALRWVQGGQGCPGVAPEVTEGAKGLEDTEPPEEEEEGEEPNYPLELLYCLY 360

Qy       361  ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       361  ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420

Qy       421  KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD 480
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       421  KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD 480
```

CC    to express a protein which affords protection against intracellular
CC    parasites and also for diagnosis of disorders associated with expression
CC    of GCREC. GCREC is also useful for generating hybridisation probes useful
CC    in mapping the naturally occurring genomic sequences and to create
CC    knockin humanised animals (pigs) or transgenic animals (mice or rats) to
CC    model human diseases. GCREC is used to diagnose, prevent and treat
CC    proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
CC    hepatitis and cancer) cancer (breast, bladder, bone marrow, brain,
CC    uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)
CC    neurologica. disorders (epilepsy, stroke, Alzheimer's, Huntington's,
CC    Parkinson's disease, multiple sclerosis, dementia and other central
CC    nervous system disorders); cardiovascular disorders (angina pectoris,
CC    hypertension, atherosclerosis, corgestive heart failure);
CC    gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal
CC    spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
CC    pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
CC    inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
CC    Addison's diseases, allergies, anaemia, asthma, diabetes mellitus, atopic
CC    dermatitis, glomerulonephritis, Grave's disease, osteoarthritis,
CC    psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
CC    parasitic, protozoal and helminthic infections) and metabolic disorders
CC    (obesity, osteoporosis, viral infections).
XX
SQ    Sequence    891 AA;

Query Match         100.0%;      Score 3218;   DB 22;    Length 891;
Best Local Similarity   100.0%;      Pred. No. 9.2e-314;
Matches   625;   Conservative    0;   Mismatches    0;   Indels    0;   Gaps    0;

Qy       1 MLAQPQRLLFILDGADELPALSGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 MLAQPQRLLFILDGADELPALSGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT  326

Qy      61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF  386

Qy     121 V?FVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFITSVLLFITSVLLFITSVADGPRLQGDLRNLC  180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 VFFVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFITSVLLFITSVADGPRLQGDLRNLC  446

Qy     181 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGV.ETEVTYQFIDQSFQEFL  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL  506

Qy     241 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLV.TRFLFGLLSAERMRDIERHFGCM  300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM  566

Qy     301 VSRRVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEQGEEPNYPLELLYCLY  360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     567 VSRRVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEQGEEPNYPLELLYCLY  626

Qy     361 E?QEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK  420
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     627 E?QEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK  686

Qy     421 KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD  480
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     687 KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD  746

Qy     481 LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPCCRVQTVRVQLPDEQRGLQYLVGX  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     747 LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPCCRVQTVRVQLPDPQRGLQYLVGX  806

Qy     541 LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRA  600
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     807 LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRA  866

Qy     601 KPDLVITHPALDGHPQPPKELISTF  625
           |||||||||||||||||||||||||
Db     867 KPDLVITHPALDGHPQPPKELISTF  891



RESULT 3
ABJ65075
ID   ABU65075 standard; Protein; 676 AA.
XX
AC   ABU65075;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Human NOV18a protein.
XX
KW   NOVX; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
KW   hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW   human.
XX
OS   Homo sapiens.
XX
PN   WC20272757-A2.
XX
PD   19-SEP-2002.
XX
PF   08-MAR-2002; 2002WO-US06908.
XX
PR   08-MAR-2001; 2001US-274101P.
PR   08-MAR-2001; 2001US-274194P.
PR   08-MAR-2001; 2001US-274281P.
PR   08-MAR-2001; 2001US-274322P.
PR   09-MAR-2001; 2001US-274849P.
PR   12-MAR-2001; 2001US-275235P.
PR   13-MAR-2001; 2001US-275578P.
PR   13-MAR-2001; 2001US-275579P.
PR   13-MAR-2001; 2001US-275601P.
PR   14-MAR-2001; 2001US-276000P.
PR   16-MAR-2001; 2001US-276776P.
PR   19-MAR-2001; 2001US-276994P.
PR   20-MAR-2001; 2001US-277239P.
PR   20-MAR-2001; 2001US-277321P.
PR   20-MAR-2001; 2001US-277327P.
PR   21-MAR-2001; 2001US-277791P.
PR   22-MAR-2001; 2001US-277833P.
PR   23-MAR-2001; 2001US-278152P.
PR   26-MAR-2001; 2001US-278894P.
PR   27-MAR-2001; 2001US-278999P.
PR   27-MAR-2001; 2001US-279036P.
PR   28-MAR-2001; 2001US-279344P.
PR   30-MAR-2001; 2001US-277338P.
PR   30-MAR-2001; 2001US-279995P.
PR   30-MAR-2001; 2001US-280233P.
PR   02-APR-2001; 2001US-280802P.
PR   02-APR-2001; 2001US-280822P.
PR   02-APR-2001; 2001US-280900P.
PR   04-APR-2001; 2001US-281194P.
PR   13-APR-2001; 2001US-283675P.
PR   30-APR-2001; 2001US-287424P.
PR   02-MAY-2001; 2001US-280066P.
PR   03-MAY-2001; 2001US-283342P.
PR   03-MAY-2001; 2001US-285528P.
PR   15-MAY-2001; 2001US-291190P.
PR   16-MAY-2001; 2001US-291099P.
PR   16-MAY-2001; 2001US-291240P.
PR   30-MAY-2001; 2001US-294485P.
PR   31-MAY-2001; 2001US-294889P.
PR   31-MAY-2001; 2001US-294899P.
PR   18-JUN-2001; 2001US-299027P.
PR   19-JUN-2001; 2001US-299303P.
PR   19-JUN-2001; 2001US-293310P.
PR   10-JUL-2001; 2001US-303354P.
PR   31-JUL-2001; 2001US-309198P.
PR   16-AUG-2001; 2001US-313903P.
PR   10-SEP-2001; 2001US-318462P.
PR   12-SEP-2001; 2001US-318770P.
PR   27-SEP-2001; 2001US-325430P.
PR   27-SEP-2001; 2001US-325681P.

```
PR   18-OCT-2001;  2001US-330380P.
PR   31-OCT-2001;  2001US-335301P.
PR   14-NOV-2001;  2001US-332172P.
PR   14-NOV-2001;  2001US-332271P.
PR   14-NOV-2001;  2001US-332272P.
PR   14-NOV-2001;  2001US-333184P.
PR   14-NOV-2001;  2001US-333272P.
PR   21-NOV-2001;  2001US-332094P.
PR   03-DEC-2001;  2001US-337426P.
PR   03-DEC-2001;  2001US-338092P.
PR   04-DEC-2001;  2001US-337185P.
PR   03-JAN-2002;  2002US-3457C5P.
PR   07-MAR-2002;  2002US-0092900.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Padigaru M,  Spytek KA,  Shenoy SG,  Taupier RJ,  Pena CEA,  Li L;
PI   Zerthuser BD,  Gusev V,  Ji W,  Gorman L,  Miller CE,  Kekuda R;
PI   Patturajan M,  Gangolli E,  Verret CAM,  Guo V,  Tcherney V;
PI   Fernandes ER,  Casman SJ,  Malyankar UM,  Gerlach V,  Liu Y;
PI   Anderson D,  Spaderna SK,  Catterton E,  Burgess C,  Leite M,  Zhong H;
PI   Alsobrock JP,  Lepley DM,  Rieger DK;
XX
DR   WPI; 2002-723332/78.
DR   N-PSDB; ABX97042.
XX
PT   NOVX polypeptides and polynucleotides, useful for preventing or
PT   treating a disorder associated with aberrant NOVX expression or
PT   activity e.g. cancer, hypertension, atherosclerosis, cardiomyopathy or
PT   bronchial asthma -
XX
PS   Claim 1; Page 149; 1103pp; English.
XX
CC   This invention describes novel human NOVX polypeptides which have
CC   been associated with a disorder associated with aberrant NOVX
CC   cyrostatic, cardiant, antiarteriosclerotic, antiasthmatic and
CC   hypotensive activity. Pharmaceutical compositions comprising the NOVX
CC   proteins or nucleic acid molecules or NOVX antibodies are useful for
CC   preventing or treating a disorder associated with aberrant NOVX
CC   expression or activity e.g. cancer, hypertension, atherosclerosis,
CC   cardiomyopathy or bronchial asthma. The products of the invention can
CC   be used for gene therapy or in a vaccine. ASU65041-ABU65218 represent
CC   the NOVX polypeptides encoded by ABX97008-ABX97185.
XX
SQ   Sequence  676 AA;

Query Match            97.7%;  Score 3144;  DB 23;  Length 676;
Best Local Similarity  98.4%;  Pred. No. 1.7e-306;
Matches  615;  Conservative  0;  Mismatches  4;  Indels  6;  Gaps  1;

Qy      1  MLAQPQRLLFILDGADELPA.GGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT  60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     58  MLAQPQRLLFILDGADELPA.GGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 117

Qy     61  RAAAPGRLQQRLCSPQCAEVRGFSKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    118  RAAAPGRLQGRLCSPQCAEVRGFSKDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 177

Qy    121  VPFVCWIVCTVLRQQLELGRCLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    178  VPFVCWIVCTVLRQQLELGRCLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC 237

Qy    181  RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    238  RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQESF 297

Qy    241  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHJVLTTRFLFGLLSAERMRDIERHFGCM 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    298  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 357

Qy    301  VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEGEEPNYPLELLYCLY 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    358  VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEGEEPNYPLELLYCLY 417
```

```
Qy    36:  ETQEDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAGQA.RLISCRLVAAQEKK 420
          |||||||||||||||||||||||||||||||||||||||||||||  ||||||||||||
Db    418  ETQEDAFVRQALGRFPELALQRVFCRMDVAVLSYCVRCCPAAQAIRL:SCRLVAAQEKK 477

Qy    421  KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD 480
          |||||||||||      |||||||||||||||||||||||||||||||||||||||||||
Db    478  KKSLGKRLQASL------GTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD 531

Qy    481  LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDQRGLQYSVGM 540
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    532  LSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDQRGLQYLVGM 591

Qy    541  LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSASVELSEQSLQELQAVKRA 600
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    592  LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRA 651

Qy    601  KPDLVITHPALDGHPQPPKELISTF 625
          |||||||||||||||||||||||||
Db    652  KPDLVITHPALDGHPQPPKELISTF 676
```

```
RESULT 4
AAU00023
ID   AAU00023 standard; Protein; 1851 AA.
XX
AC   AAU00023;
XX
DT   10-MAY-2001 (first entry)
XX
DE   Human activated T-lymphocyte associated sequence 2, ATLAS-2.
XX
KW   Human; activated T-lymphocyte associated sequence 2; ATLAS-2; antibody;
KW   cytokine receptor; autoimmune disorder; immune disorder; cancer;
KW   T-lymphocyte-associated disorder; cell proliferation disorder; tumour;
KW   cell differentiation disorder; immune deficiency disorder; malignancy;
KW   viral infection; bacterial infection; fungal infection; metabolism;
KW   chromosome 11p15.5.
XX
OS   Homo sapiens.
XX
FN   WO20011464-A2.
XX
PD   01-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-US22699.
XX
PR   20-AUG-1999; 99US-0150105.
PR   28-APR-2000; 2000US-0560101.
PR   28-APR-2000; 2000US-0560365.
PR   28-APR-2000; 2000US-0560948.
PR   28-APR-2000; 2000US-0561533.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOG ) BIOGEN INC.
XX
PI   Peyman JA,  Green CD,  Hsu A,  Browning JA,  Carulli J;
XX
DR   WPI; 2001-218453/22.
DR   N-PSDB; AAS00033.
XX
PT   New isolated activated T lymphocyte associated sequences for treating
PT   or preventing immune system associated disorders such as autoimmune
PT   disorder, immune disorder, and T-lymphocyte-associated disorder -
XX
PS   Claim 14; Fig 2; 114pp; English.
XX
CC   The sequence represents human activated T-lymphocyte associated sequence
CC   2, ATLAS-2. ATLAS-2 is related by homology to cytokine receptors and its
CC   gene is located on chromosome 11p15.5. ATLAS proteins, polynucleotides
CC   and antibodies are useful for treating/preventing conditions associated
CC   with an autoimmune disorder, immune disorder, T-lymphocyte-associated
CC   disorder, cell proliferation disorder, cell differentiation disorder,
CC   and immune deficiency disorder and for screening for a modulator of
```

CC　　activity or of latency or predisposition to an immune disorder.
CC　　ATLAS proteins, polynucleotides and antibodies are useful in therapeutic
CC　　or prophylactic treatment of diseases associated with cell proliferation
CC　　(e.g. cancers, malignancies and tumours). The polynucleotides are useful
CC　　in gene therapy, to detect ATLAS mRNA or a genetic lesion in an ATLAS
CC　　gene, to modulate ATLAS activity, to screen drugs or compounds that
CC　　modulate ATLAS activity or expression and to treat disorders
CC　　characterised by insufficient or excessive production of ATLAS protein or
CC　　production of ATLAS protein forms that have decreased or aberrant
CC　　activity compared to ATLAS wild type protein and in tissue typing to
CC　　identify individuals. The antibodies are useful for localisation/
CC　　quantitation, isolation and detection of ATLAS and to monitor protein
CC　　levels in tissue. ATLAS is useful for treating/preventing infection by
CC　　bacteria, viruses and fungi, affecting bodily characteristics, e.g.
CC　　biorhythms, fertility or metabolism, affecting behavioural
CC　　characteristics, and for providing analgesic effects. A host cell
CC　　containing the polynucleotide is useful to produce non-human transgenic
CC　　animals.
XX
SQ　　Sequence 1851 AA;

Query Match　　　89.6%;　　Score 2882;　　DB 22;　　Length 1851;
Best Local Similarity　　81.8%;　　Pred. No. 1.7e-279;
Matches 581;　　Conservative 7;　Mismatches 32;　　Indels 90;　Gaps 5;

Qy　　　1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　267 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 326

Qy　　 61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　327 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 386

Qy　　121 VPFVCWIVCTVLRQQLEGRDLSRTSKTTTSVLLFITSVLSSAPVADGPRLQGDLRNLC 180
　　　　　|||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　387 VPFVCWIVCTVLRQQLEGRDLSRTSKTTTSVLLFITSVLSSAPVADGPRLQGDLRNLC 445

Qy　　181 RLAREGVLGRRAQFAEKELEQLSLRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 240
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　447 RLAREGVLGRRAQFAEKELEQLSLRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 506

Qy　　241 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 300
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　507 AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMSDIERHFGCM 566

Qy　　301 VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY 360
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　567 VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY 626

Qy　　361 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　627 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 686

Qy　　421 KKSLGKRLQASLGGSSSQT------------------- 440
　　　　　||||||||||||||||||||
Db　　687 KKSLGKRLQASLGGSWLGTQLAPEVFRPPCCDICPTPPPDPRLLQGKAFARVPLNIAP 746

Qy　　441 TKQLPASLL--------- 484
　　　　　|||||||||
Db　　747 IQPLPRGLASVERMXNVTVLAGAGPGDPKTHAMTDPLCHLSSLTLSHCKLPDAVCRDLSEA 806

Qy　　485 LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYIVGMLRQS 544
　　　　　||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db　　807 LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYIVGMLRQS 866

Qy　　545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLA--------SVELSE------ 588
　　　　　|||||||||||||||||||||||||||||||||||||        |||||
Db　　867 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLSLPSDPTPSSFSGRCREPGRRLG 926

Qy　　589 ------QSLQELQAVKRAKPDLVIT--------HPALDGHPQPPKE 620
　　　　　　　　　　　　　　　　　　　　　　　|||||||||||
Db　　927 LESRWPRSAPEPSCDSEARTQVEAAGGAGGGRRGREPPARGPHPQPPRD 976

```
XX
DR   WPI; 2001-457740/49.
DR   N-PSDB; ABA09979.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 20; Page 236; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, stem cell growth factor
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medical
CC   conditions, e.g. by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g., myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a novel human
CC   polypeptide of the invention.
XX
SQ   Sequence  597 AA;

Query Match     69.6%;   Score 2239;   DB 22;   Length 597;
Best Local Similarity  100.0%;   Pred. No. 1.3e-215;
Matches 435;   Conservative   0;   Mismatches   0;   Indels   0;   Gaps   0;

Qy    1  MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT   60
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  163  MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT  222

Qy   61  RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAEBAYRFVKENETLFALCF  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  223  RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAEBAYRFVKENETLFALCF  282

Qy  121  VPFVCWIVCTVLRQQLEIGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNLC   180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  283  VPFVCWIVCTVLRQQLEIGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNLC   342

Qy  181  RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL  240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  343  RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL  402

Qy  241  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM   300
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  403  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM   462
```

```
Db  447  RLAREGVLGRRAQFAEKEDQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL  506
Qy  241  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM   300
Db  507  AALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM  566

Qy  301  VSERVKQEALRWVQQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY   360
Db  567  VSERVKQEALRWVQQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY   626

Qy  361  ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK  420
Db  627  ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK  686

Qy  421  KKSLGKRLQASLGGGSSQGT-----------------------------------  440
Db  687  KKSLGKRLQASLGGSWLGTQLAPEVPVFRPPCCDICPTPPPDPRLLQGKAFARVPLNIAP  746

Qy  441  TKQLPASLL--------------HPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEA  484
Db  747  ICPLPRGLASVERMVTVLAJAGPGDPKTHAMTDPLCHLSSLTLSHCKLPDAVCRDLSEA  806

Qy  485  LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQS  544
Db  807  LRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQS  866

Qy  545  PALTLDLSGCQLPAPMVTYLCAVLQHGGCGLQTLSLA------------SVELSE------  588
Db  867  PALTLDLSGCQLPAPMVTYLCAVLQHGGCGLQTLSLSLPSDPTPSSFSGRCREPGRRLG  926

Qy  589  -------QSLQELQAVXRAKPDLVIT----------HPALDGHPQPPKE  620
Db  927  LESRWPRSAPEPSGDSEARTQVEAAGGAGGGRRGREPPARGPHPQPPRD  976
```

```
RESULT 6
ABB11735
ID   ABB11735 standard; peptide; 597 AA.
XX
AC   ABB11735;
XX
DT   11-JAN-2002 (first entry)
XX
DE   Human vasopressin receptor homologue, SEQ ID NO:2105.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer.
XX
OS   Homo sapiens.
XX
PN   WO200157188-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
```

```
CC   receptors respond to arginine vasopressin. Isoreceptors have been
CC   described for both AVP and AII, this is consistent with their
CC   functional diversity. This recombinant receptor may be derived from
CC   a mammal, preferably a human or a rat. The DNA encoding this sequence
CC   may be used in the construction of vectors for the expression of the
CC   recombinant receptor. Fragments of the extracellular domain of the
CC   receptor are useful as immunogens for producing antibodies which
CC   neutralise the activity of the AII/AVPv2 receptor in vivo.
XX   (Updated on 25-MAR-2003 to correct PN field.)
SQ   Sequence 481 AA;

Query Match             49.0%;  Score 1578;  DB 14;  Length 481;
Best Local Similarity   67.1%;  Pred. No. 2.7e-149;
Matches  328;  Conservative  49;  Mismatches  98;  Indels  14;  Gaps  5;

Qy    136 LELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFA 195
              |||||||||||||||||||| ::|    | ||  | |||| || ||||||||: ||| |
Db      1 MELGRDLSRTSKTTTSVYLFITSMLKSAG-TNGPRVQGELRMLCRLAREGILKHQAQFS 59

Qy    196 EKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFLAALSYLLEDGGVPRT 255
              || || ::|:| ||| ||||::|||||||.| |||||||||||||||||||||  ||  |
Db     60 EKDLERLKLQGSQVQTMFLSKNQLPGVLETVVTYQFIDQSFQEFLAALSYLLDAEGAPGN 119

Qy    256 AAGGVGTLLRGDAQPHSHLVLTTRFLFGGLLSAERMRDIERHFGCMVSERVKQEALRWVQG 315
              ::| | ||| |   ::| |||||||||| |||:||:||||:| |||| :|||| |::||||
Db    120 SAGSVQMLLLSDAGLRGHLALTTRFLFGLLSTERIRDIGNHFGCVVPGRVKQDTLRWVQG 179

Qy    316 QGQCPGVA---PEVTEGAKGLEDTEEPEEEGEEPNYPLELYCLYETQEDAFVRQAL 372
              | |   || :::|                             ::::::::::::::
Db    180 QSQ--PKVATVGAEKKDELKDEEAEEEEEEEEELNFGLELYCLYETQEDDFVRQAL 237

Qy    373 CRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQE--KKKKSLGKRLQA 430
              .: ||:.|:|||::|||:: :|||| |||.|||||:||. |||:||    ||| ::||:
Db    238 SSLPEMVLERVRLTRMDLEVLSYCVQCCPGQALRLVSCGLVAAKEKKKKKSFMNRLK- 296

Qy    431 SLGGGSSQGTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPA 490
               ::|:::|: :: ||||| ||  ||  .. | || ||||||||||||||||||| ::||:
Db    297 ----GSQSTGKQPPASLLRPLCEAMITQQCGLSILTLSHCKLPDAVCRDLSEALKVAPS 351

Qy    491 LTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQSPALTL 550
              | |||||:|||:||||| |:|||||| |:|:|| :: ::: :: : :| :| :|| | | 
Db    352 LRELGLLQNRLTEAGLRLLSQGLAWPKCKVQTLRIQMPGLQEVIHYLVLVLQQSPVLTTL 411

Qy    551 DLSGGQLPAPMVTYLCAVLQHQCGCGLQTLSLASVELSEQSLQEIQAVKRAKPDLVTHPA 610
              ||||||||. .| . ||:|:| | ||::||| ::|||| : |: ||:| | | ||| :| 
Db    412 DLSGGQLPGTVVEPLCSALKHPKCGLKTLSLTSVELTENPLRELQAVKTLKPDLAIIHSK 471

Qy    611 LDGHPQPPK 619
              | ||||| |
Db    472 LGTHPQPLK 480

RESULT 8
ABB90173
ID   ABB90173 standard; Protein; 321 AA.
XX
AC   ABB90173;
XX
DT   24-MAY-2002 (first entry)
XX
DE   Human polypeptide SEQ ID NO 2549.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein.
XX
OS   Homo sapiens.
XX
PN   WO200290304-A2.
XX
```

```
Qy    301 VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELYCLY 360
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    463 VSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELYCLY 522

Qy    361 ETQEDAFVRQALCRFPELALCRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    523 ETQEDAFVRQALCRFPELALCRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 582

Qy    421 KKSLGKRLQASLGGG 435
              |||||||||||||||
Db    583 KKSLGKRLQASLGGG 597

RESULT 7
AAR33389
ID   AAR33389 standard; Protein; 481 AA.
XX
AC   AAR33389;
XX
DT   25-MAR-2003 (updated)
DT   15-JUL-1993 (first entry)
XX
DE   AII/AVPv2 receptor.
XX
KW   Recombinant; angiotensin ii cAMP; vasopressin-v2; AII; AVPv2; vector;
KW   receptor; G protein-coupled; neurotransmitter; angiotensin ii; human;
KW   arginine vasopressin; isoreceptor; mammal; rat; extracellular domain;
KW   immunogen; antibody; AII/AVPv2.
XX
OS   Synthetic.
XX
FH   Key             Location/Qualifiers
FT   Peptide         392..399
FT                   /note= "Antagonist peptide"
FT   Peptide         342..350
FT                   /note= "Antagonist peptide"
FT   Domain          30..94
FT                   /note= "Extracellular domain, immunogenic peptide"
FT   Domain          151..251
FT                   /note= "Extracellular domain, immunogenic peptide"
FT   Domain          193..200
FT                   /note= "Pref. immunogenic peptide"
FT   Domain          338..390
FT                   /note= "Extracellular domain, immunogenic peptide"
FT   Domain          437..481
FT                   /note= "Extracellular domain, immunogenic peptide"
XX
PN   WO9305073-A1.
XX
PD   18-MAR-1993.
XX
PF   11-SEP-1992; 92WO-US07786.
XX
PR   11-SEP-1991; 91US-0758921.
XX
PA   (UYBO-) UNIV BOSTON.
XX
PI   Herrera VLM, Ruiz-Opazo N;
XX
DR   WPI; 1993-100926/:2.
DR   N-PSDB; AAQ38723.
XX
PT   Recombinant angiotensin II cAMP/vasopressin V2 receptors and
PT   their analogues - for treating and preventing hypertension and
PT   stroke, and for diagnosing hypertension
XX
PS   Claim 2; Fig 1; 7ipp; English.
XX
CC   This sequence represents a recombinant angiotensin II cAMP/vasopressin-
CC   v2 (AII/AVPv2) receptor polypeptide. AII and AVP receptors are both
CC   G protein-coupled receptors with diverse physiological roles. AII
CC   receptors respond to the neurotransmitter angiotensin II whilst AVP
```

PD   29-NOV-2001.
XX
PF   18-MAY-2001; 2001:WO-US16450.
XX
PR   19-MAY-2000; 2000US-205515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE,  Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   N-PSDB; ABL90582.
XX
PT   Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 11; SEQ ID NO 2549; 2087pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence  321 AA;

Query Match      34.8%;  Score 1120;  DB 23;  Length 321;
Best Local Similarity  86.5%;  Pred. No. 1.6e-103;
Matches 230;  Conservative   5;  Mismatches  22;  Indels  10;  Gaps   4;

Qy     1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT  60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    58 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 117

Qy    61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFA-CF 129
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||| ||
Db   118 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFXFRDERRAERAYRFVKENETLFA-CF 177

Qy   122 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNLC 180
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   178 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNLC 237

Qy   181 RLAREGVLGRRAQFAEKELEQLELRGSK-VQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
           ||||||||||||  ||||||||||||| || |||||||||||||||||||||||||| |
Db   238 RLAREGVLGRKAQFAXKELEQLELRGSKSADACFSAKRSCPGVLXTEVTYQFID---QEL 294

Qy   240 LAALSYLL-----EDG-GVPRTAAGG 259
           |      ||    ||| ||||||||
Db   295 XRSSSRALVLPCWKDGRGGFRNRRXG 320

RESULT 9
AAM89960
ID   AAM89960 standard; Protein; 329 AA.
XX
AC   AAM89960;
XX
DT   07-NOV-2001  (first entry)

XX
DE   Human immune/haematopoietic antigen SEQ ID NO:17553.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000;  2000US-0179065.
PR   04-FEB-2000;  2000US-0180628.
PR   24-FEB-2000;  2000US-0184664.
PR   02-MAR-2000;  2000US-0186350.
PR   16-MAR-2000;  2000US-0189874.
PR   17-MAR-2000;  2000US-0190076.
PR   18-APR-2000;  2000US-0198123.
PR   19-MAY-2000;  2000US-0205515.
PR   07-JUN-2000;  2000US-0209467.
PR   28-JUN-2000;  2000US-0214886.
PR   30-JUN-2000;  2000US-0215135.
PR   07-JUL-2000;  2000US-0216647.
PR   07-JUL-2000;  2000US-0216880.
PR   11-JUL-2000;  2000US-0217487.
PR   11-JUL-2000;  2000US-0217496.
PR   14-JUL-2000;  2000US-0218290.
PR   26-JUL-2000;  2000US-0220963.
PR   26-JUL-2000;  2000US-0220964.
PR   14-AUG-2000;  2000US-0224518.
PR   14-AUG-2000;  2000US-0224519.
PR   14-AUG-2000;  2000US-0225213.
PR   14-AUG-2000;  2000US-0225214.
PR   14-AUG-2000;  2000US-0225266.
PR   14-AUG-2000;  2000US-0225267.
PR   14-AUG-2000;  2000US-0225268.
PR   14-AUG-2000;  2000US-0225270.
PR   14-AUG-2000;  2000US-0225447.
PR   14-AUG-2000;  2000US-0225757.
PR   14-AUG-2000;  2000US-0225758.
PR   14-AUG-2000;  2000US-0225759.
PR   18-AUG-2000;  2000US-0226279.
PR   22-AUG-2000;  2000US-0226681.
PR   22-AUG-2000;  2000US-0226868.
PR   22-AUG-2000;  2000US-0227182.
PR   23-AUG-2000;  2000US-0227009.
PR   30-AUG-2000;  2000US-0228924.
PR   01-SEP-2000;  2000US-0229287.
PR   01-SEP-2000;  2000US-0229341.
PR   01-SEP-2000;  2000US-0229344.
PR   01-SEP-2000;  2000US-0229345.
PR   05-SEP-2000;  2000US-0229509.
PR   05-SEP-2000;  2000US-0229513.
PR   06-SEP-2000;  2000US-0230437.
PR   06-SEP-2000;  2000US-0230438.
PR   08-SEP-2000;  2000US-0231242.
PR   08-SEP-2000;  2000US-0231243.
PR   08-SEP-2000;  2000US-0231244.
PR   08-SEP-2000;  2000US-0231413.
PR   08-SEP-2000;  2000US-0231414.
PR   08-SEP-2000;  2000US-0232080.
PR   08-SEP-2000;  2000US-0232081.
PR   12-SEP-2000;  2000US-0231968.
PR   14-SEP-2000;  2000US-0232397.
PR   14-SEP-2000;  2000US-0232398.
PR   14-SEP-2000;  2000US-0232399.
PR   14-SEP-2000;  2000US-0232400.
PR   14-SEP-2000;  2000US-0232401.
PR   14-SEP-2000;  2000US-0233063.
PR   14-SEP-2000;  2000US-0233064.

```
PR   14-SEP-2000; 2000US-0233065.
PR   21-SEP-2000; 2000US-0234223.
PR   25-SEP-2000; 2000US-0234274.
PR   25-SEP-2000; 2000US-0234997.
PR   25-SEP-2000; 2000US-0234998.
PR   26-SEP-2000; 2000US-0235484.
PR   27-SEP-2000; 2000US-0235834.
PR   27-SEP-2000; 2000US-0235836.
PR   29-SEP-2000; 2000US-0236327.
PR   29-SEP-2000; 2000US-0236367.
PR   29-SEP-2000; 2000US-0236368.
PR   29-SEP-2000; 2000US-0236369.
PR   29-SEP-2000; 2000US-0236370.
PR   02-OCT-2000; 2000US-0236802.
PR   02-OCT-2000; 2000US-0237037.
PR   02-OCT-2000; 2000US-0237038.
PR   02-OCT-2000; 2000US-0237039.
PR   02-OCT-2000; 2000US-0237040.
PR   13-OCT-2000; 2000US-0239935.
PR   13-OCT-2000; 2000US-0239937.
PR   20-OCT-2000; 2000US-0240960.
PR   20-OCT-2000; 2000US-0241221.
PR   20-OCT-2000; 2000US-0241785.
PR   20-OCT-2000; 2000US-0241786.
PR   20-OCT-2000; 2000US-0241787.
PR   20-OCT-2000; 2000US-0241808.
PR   20-OCT-2000; 2000US-0241809.
PR   20-OCT-2000; 2000US-0241826.
PR   01-NOV-2000; 2000US-0244617.
PR   01-NOV-2000; 2000US-0244474.
PR   08-NOV-2000; 2000US-0246475.
PR   08-NOV-2000; 2000US-0246476.
PR   08-NOV-2000; 2000US-0246477.
PR   08-NOV-2000; 2000US-0246478.
PR   08-NOV-2000; 2000US-0246523.
PR   08-NOV-2000; 2000US-0246524.
PR   08-NOV-2000; 2000US-0246525.
PR   08-NOV-2000; 2000US-0246526.
PR   08-NOV-2000; 2000US-0246527.
PR   08-NOV-2000; 2000US-0246528.
PR   08-NOV-2000; 2000US-0246532.
PR   08-NOV-2000; 2000US-0246609.
PR   08-NOV-2000; 2000US-0246610.
PR   08-NOV-2000; 2000US-0246611.
PR   08-NOV-2000; 2000US-0246613.
PR   17-NOV-2000; 2000US-0249207.
PR   17-NOV-2000; 2000US-0249208.
PR   17-NOV-2000; 2000US-0249209.
PR   17-NOV-2000; 2000US-0249210.
PR   17-NOV-2000; 2000US-0249211.
PR   17-NOV-2000; 2000US-0249212.
PR   17-NOV-2000; 2000US-0249213.
PR   17-NOV-2000; 2000US-0249214.
PR   17-NOV-2000; 2000US-0249215.
PR   17-NOV-2000; 2000US-0249216.
PR   17-NOV-2000; 2000US-0249217.
PR   17-NOV-2000; 2000US-0249244.
PR   17-NOV-2000; 2000US-0249245.
PR   17-NOV-2000; 2000US-0249264.
PR   17-NOV-2000; 2000US-0249265.
PR   17-NOV-2000; 2000US-0249297.
PR   17-NOV-2000; 2000US-0249299.
PR   17-NOV-2000; 2000US-0249300.
PR   01-DEC-2000; 2000US-0250160.
PR   01-DEC-2000; 2000US-0250391.
PR   05-DEC-2000; 2000US-0251030.
PR   05-DEC-2000; 2000US-0251988.
PR   05-DEC-2000; 2000US-0256719.
PR   06-DEC-2000; 2000US-0251479.
PR   08-DEC-2000; 2000US-0251856.
PR   08-DEC-2000; 2000US-0251868.
PR   08-DEC-2000; 2000US-0251869.
```

```
PR   08-DEC-2000; 2000US-0251989.
PR   08-DEC-2000; 2000US-0251990.
PR   11-DEC-2000; 2000US-0254097.
PR   05-JAN-2001; 2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-4B3426/52.
DR   N-PSDB; AAK62741.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 11; SEQ ID NO 17553; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence  329 AA;

Query Match             32.8%;  Score 1056.5;  DB 22;  Length 329;
Best Local Similarity   83.9%;  Pred. No. 4.1e-97;
Matches  219;  Conservative    5;  Mismatches  34;  Indels    3;  Gaps    3;

Qy           1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
               |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db          66 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 125

Qy          61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         126 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKYFYKFFRDERRAERAYRFVKENETLFALCF 185

Qy         121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLLSSAPVADGPRLQGDLRNLC 180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db         186 VPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLLFITSVLLSSAPVADGPRLQGDLRNLC 245

Qy         181 RLAREGVLGRRAQFAEKELEQLELRGSK-VQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
               |||||||||| ||||||||||||||||   |:|
Db         246 RLAREGVLGRXAQFAEKELEQLELRGSSKSADACFSAKRXCRACWXQRSPTSSSTKSFXGV 305

Qy         240 -LAALSYLL-EDGGVPRTAAG 258
                ||||||| ||  :  |:|
Db         306 PRGHLSYLAGRTGGVAQGTAG 326

RESULT 10
AAO15590
ID   AAO15590 standard; Protein; 1061 AA.
XX
AC   AAO15590;
XX
DT   31-OCT-2002  (first entry)
XX
DE   Human PYRIN-8 protein #2.
XX
```

```
KW   Human; gene therapy; PYRIN; stress-related response; apoptotic response;
KW   inflammatory response; inflammatory disorder; immune system disorder;
KW   Crohn's disease; multiple sclerosis; cancer; leukaemia;
KW   autoimmune disorder; arthritis; neurological disease;
KW   Alzheimer's disease; Parkinson's disease; chromosomal mapping;
KW   tissue typing; forensic biology; predictive medicine; pharmacogenomics;
KW   transcription profiling; PYRIN-8.
XX
OS   Homo sapiens.
XX
PN   WO20026-049-A2.
XX
PD   08-AUG-2002.
XX
PF   31-JAN-2002; 2002WO-US02967.
XX
PR   31-JAN-2001; 2001US-2652231P.
PR   10-SEP-2001; 2001US-318645P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (AMHP ) WYETH.
XX
PI   Bertin J, Wang W, Blatcher M;
XX
DR   WPI; 2002-627477/67.
DR   N-PSDB; AAL44363.
XX
PT   New PYRIN polypeptides and nucleic acids useful for modulating and
PT   diagnosing stress-related, apoptopic and inflammatory responses, or for
PT   treating inflammatory and immune system disorders, cancers, or
PT   neurological diseases -
XX
PS   Claim 8; Fig 8; 167pp; English.
XX
CC   The invention comprises the amino acid and DNA coding sequences of human
CC   PYRIN proteins. The PYRIN protein and DNA sequences of the invention are
CC   useful for modulating and diagnosing stress-related, apoptopic and
CC   inflammatory responses. The PYRIN protein and DNA sequences are useful
CC   for treating inflammatory disorders and immune system disorders (e.g.
CC   Crohn's disease, reactive arthritis, multiple sclerosis, contact
CC   dermatitis, psoriasis, graft rejection, allergies, viral infections and
CC   bacterial infections), cancer (e.g. leukaemia); autoimmune disorders
CC   (e.g. systemic lupus erythematosus and arthritis); and neurological
CC   diseases (e.g. Alzheimer's disease and Parkinson's disease). The PYRIN
CC   protein and DNA sequences may also be used in screening assays, detection
CC   assays (e.g. chromosomal mapping, tissue typing or forensic biology),
CC   predictive medicine (e.g. diagnostic assays, clinical trials and
CC   pharmacogenomics) and transcription profiling. The present amino acid
CC   sequence represents a human PYRIN-8 protein.
XX
SQ   Sequence 1061 AA;

Query Match        25.7%;  Score 825.5;  DB 23;  Length 1061;
Best Local Similarity  32.6%;  Pred. No. 4.3e-73;
Matches 219;  Conservative 104;  Mismatches 245;  Indels 103;  Gaps 15;

Qy      1 MLAQPQRLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLFTALLLVT 59
            |: |||:::||| ||||| ||: :|||:||:: ::||||||| |||| ||
Db    282 LIRVPERLLFIIDGFDELXPSFHDPQGPWCLCWEEKRPTELLNSLIRKKLLPELSLLT 341

Qy     60 TRAAAPGRLQGRLCSPQCAE?RGFSDKDKKKYFKFFRDERRAERAYRFVKENETLFALC 119
            |:: |: ||: : ::|| :| || ||||:|||:  : ||| ||::|:  |::||||
Db    342 TRPTALEKLHRLLEHPRHVEILGFSEAERKEYFYKYFHNAECASQVFNVVRDNEPLFTMC 401

Qy    120 FV?FVCWIVCTVLRQQLE.GRDLSRTSKTT?SVYLLFITSVLSSAPVADGPRLQ--GDLR 177
            |  |:|: |:| :|||:| :|::||:|: || :|:  |::||:|||: :|::
Db    402 FVPLVCWVVCTCLQQQLEGGGLLRQTSRTT?AVYXLYLLSLMQPXPGA--PRLQPPPNQR 459

Qy    178 NLCRLAREGVLGRRAQFAEKLEQELERGSKVQTLFLSKKELPGVLETEVTYQFEDQSFQ 237
            |: ||: :|| : |::|:|:||::|:|:||:|:::| ::|::||:|::||::||:
Db    460 GLCSLAADGLWNQKILFEEQLRKHGLDGEDV-SAFLNMNIFQKDINCERYYSFIHLSFQ 518

Qy    238 EFLAALSYLLEDGGVPRTAAGGVG------TLLRGDAQPHSHLVLTTRFLFGLLSAERM 290
            :|||| |  :||:|::  : ::||       :||||: | :|: :| ||::|:||
```

RESULT 11
AAO17857
ID   AAO17857 standard; Protein; 1099 AA.
XX
AC   AAO17857;
XX
DT   20-AUG-2002 (first entry)
XX
DE   Pyrin domain containing protein NALP3/PYS-hs.
XX
KW   Pyrin domain; PYD domain; antiinflammatory; antiparkinsonian;
KW   antiarteriosclerotic; antipsoriatic; antibacterial; antiasthmatic;
KW   neuroprotective; antiarthritic; antirheumatic; antiasthmatic;
KW   nephrotropic; osteopathic; nootropic; intracellular signal transduction;
KW   inflammation; Alzheimer's disease; infection; psoriasis; asthma;
KW   arteriosclerosis; multiple sclerosis; rheumatoid arthritis; sarcoidosis;
KW   osteoarthritis; glomerulonephritis.
XX
OS   Unidentified.
XX
PN   WO2C0240668-A2.
XX
PD   23-MAY-2002.
XX
PF   30-OCT-2001; 2001WO-EP12545.
XX
PR   15-NOV-2000; 2000DE-1056687.
PR   30-NOV-2000; 2000DE-1059595.
XX
PA   (APCT-) APOTECH RES & DEV LTD.
XX
PI   Tschopp J, Martinon F;
XX
DR   WPI; 2002-427093/45.
DR   N-PSDB; AAL47129.
XX
PT   New DNA encoding protein with pyrin domain, useful for treating
PT   diseases involving impaired signal transduction, particularly
PT   inflammation, also proteins and antibodies -
XX
PS   Claim 5; Fig 1; 116pp; German.

```
Db    519 EFFAAWYYILDEG------------------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLEGLLNEETR 572
                                            ||||| |:|::  || ||:|||||
Qy    291 RDIER?FGCMVSERVKQEALRWVCGQQGCPGVAPEVTEGAKGLEDTEEPEEEGEEPN 350
           |:|:   ::|||:||:| |  :|| ::|:|| :|||
Db    573 SHLEKSLCWKVSPHIKMDLLQWIQSKAQ------------------SDGSTLQQGS-- 613

Qy    351 YPLELLYCLYETQEDAFVRQALCRPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLIS 410
           |: ||:|:   :  :|| ::||   ||| |::| | : : || | |:| :|:
Db    611 --LEFFSCLYEIQEEEFIQQALSHFQVIVVSNIA-SKMEH?VSSFCLKRCRSAQVLFLYG 667

Qy    411 CRLVAAQEKKKKSLGKRLQASLGGGS-----------SQGTTKQLPAS:----------- 448
           |::| |:|:|:::|  :|:||::||||            :| :||:||:
Db    668 ATYSADGE--------DRARCSAGAHTLLVQLPERTVLLDAYSEHLAAALCTNPNLIELSL 720

Qy    449 -----------LRFLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLHN 499
                       |:|| |  :| ||: ::||:||:||  |::|:::::| |: ||:|
Db    721 YRNALGSRGVKLLCQGLRHPNCKLQNLRLKRCRISSSACEDLSAALIANKNLTRMDLSGN 780

Qy    500 RLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTTLDLSGCQLP 558
           || |||||:|| ||  |||:|: :|  |:|::||  |||:|:||:| | |:||:|| ||
Db    781 GVGFPGMMLLCEGLRHFQCRLQMIQLRKCQLESGACQEMASVLGTNPHLVELDLTGNALE 840

Qy    559 APXVTYLCAVLQHQGCLQTL------------------SLASVELSEQSLQELQAVKRAKPDLV 605
           |  ::||||:|:|:|| ||:                  ||||: :|:|:| |||:|:|:|:
Db    841 DLGLRLLCQGLRHPVCR.RTLWLKICRLTAAACDELASTLSVNQSLRELDLS.NELGDLG 900

Qy    606 I------THP 609
           |       |:|
Db    90: VLLCEGLRHP 911
```

```
XX
CC   The present invention relates the DNA and their encoded proteins, where
CC   the proteins contain at least one PYD (pyrin) domain. These can be used
CC   to treat diseases associated with impaired intracellular signal
CC   transduction, particularly inflammation such as psoriasis,
CC   arteriosclerosis, bacterial or viral infections (particularly meningitis
CC   and pneumonia), multiple sclerosis, rheumatoid arthritis, asthma,
CC   sarcoidosis, glomerulonephritis and osteoarthritis, and also Alzheimer's
CC   and Parkinson's diseases. The present sequence is a protein of the
CC   invention.
XX
SQ   Sequence    1099 AA;

Query Match     25.6%;   Score 825;   DB 23;   Length 1099;
Best Local Similarity   32.6%;   Pred. No. 5.1e-73;
Matches 249;  Conservative 104;  Mismatches 245;  Indels 104;  Gaps  15;

Qy     1  MLAQPQRLFLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALL.VT  59
             ::.:  :.  :.:||.||:.:  :|:::  :||| :  ||:|:|:|| |: |:|:.|:::
Db   255  LIRVPERLLFIIDGFDELKPSFHDPQGPWCLCWEEKRPTELLLNSLIRKKLLPELSLLIT  314

Qy    60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKDKKYFYKPFRDERRAERAYRFVKENETLFALC  119
             |: :|:| :|| |:::| |: |:||::||:|::: || | :|  :|||:|:| :|::| |:
Db   315  TRPTALEKLHRLLEHPRHVEILGFSEAERKEYFYKYTHNAEQAGQVFNYVRONEPLFTMC  374

Qy   120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFTTSVLLSSAPVADGPPLQ--GDLR  177
             |||.|:|||:| |:::|:||::| | | ||.:::|:|:|:::.::|::..|::.   |:||
Db   375  FVPLVCWVVCTCLQQQLEGGGLLRQTSRTTTAVYMLILLS:NQPKPGA--PRLQPPPNQR  432

Qy   178  NLCRLAREGVLGSRAQFAEKELEQLELRGSKVCTLFLSKKELPGVLETEVVYCFIDCQSFQ  237
             |:| |:.||.:| :|  |: : ||:|:|| :|: :|:|::|||::||||:|:|:||:|:||
Db   433  GLCSLAADGLWNQKILFEEQDLRKHGLDGEDV-SAFLNMNIFQKDINCERYYSFIHLSFQ  491

Qy   238  EFLAALSYLLEDGGVPRTAAGGVG-------TLLRGDAQPNSHLVLTTRFLFGLLSAERX  290
             |: ::|.:|  ::.: |  .||.         :: :| ||||: |:.| |:|||:.||
Db   492  EFFAAMYYILDEG-------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLFGLLNEETR  545

Qy   291  RDIERHFGCMVSERVKQEALRWTVQGQGQGCPGVAPEVTEGAKGLEETEEPEEEEGEEPN  350
             ::::  |::   ::::: |:  |:  :  ::                           |:
Db   546  SHLEKSLCWKVSPHIKMDLLQWIQSKAQ-----------------------SDGSTLQQGS  583

Qy   351  YPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLIS  410
             ||: |||:::   | ::|:||: |||: |:.|::|| :|: :| :|  ||.|:.|:|| :
Db   584  --LEFFSCLYEIQEEFIQQALSHFQVIVVSNIA-SKMEHMVSSFCLKRCRSAQVLHLYG  640

Qy   411  CRLVAAQEKKKKSLGKRLQASLGGSS----------SQGTTKQLPASL----------  448
             |:::|:::  :|:|||:|:||||.||           |:|||: | ||
Db   641  ATYSADGE------DRARCSAGAHTLLVQLRPERTVLLDAYSEHLAAALCTNPNLLELS  693

Qy   449  ----------LHPLFQAMTDPLCHLSSLJTLSHCKLPDCAVCRDLSEALRAAPALTEGLLH  498
                       |:|:|:|:|| :||::::::: |:||  |::|| ::|:|::|||:|.:||
Db   694  LYRNALGSRGVKLLCQGLRHPNCKLQNLRLKRCRISSSACEDLSAALIANKNLTRXDLSG  753

Qy   499  NRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTTLDLSGCQL  557
             |::| :|:|:|::|| | || : :|::  : |::   | |:|:|:|:|:|:|:|||:|||
Db   754  NGVGFPGMMLLCEGLRHPQCRLQMIQLRKCQOLESGACQEMASVLGTNPHLVELDLTGNAL  813

Qy   558  PAPMVTYLCAVLQHQGCGLQT,-----------SLASVELSEQSLQELQAVKRAKPDL  604
             |:|:|:|||: |:|:||.||.             :|:::::|:|::||:|:|::|:|:|
Db   814  EDLGLRLLCQGLRHPVCRLRTLWKICRLTAAACDELASTLSVNQSLRELDLSLNELGEL  873

Qy   605  VI------THP  609
               |      |||
Db   874  GVLLLCEGLRHP  885
```

```
XX
DE   Human PYRIN-1 protein.
XX
KW   Human; nucleotide binding site; NBS-1; PYRIN-1; apoptotic; cytostatic;
KW   dermatological; vasotropic; melanoma; gene therapy; opthalmological;
KW   cancer; leukaemia; carcinoma; arthritis; viral infection; allergy;
KW   autoimmune disease; systemic lupus erythematosus; SLE; nootropic; asthma;
KW   glomerulonephritis; neurological disorder; Alzheimer's disease; ALS;
KW   amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
KW   aplastic anaemia; ischaemia; meningitis; liver disease; Crohn's disease;
KW   insulin-dependent diabetes; multiple sclerosis; Grave's disease; HIV;
KW   human immunodeficiency virus; tuberculosis; lepromatous leprosy.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   Domain          1..87
FT                   /label= Pyrin_domain
FT   Modified-site   3..5
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   44..46
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   93..98
FT                   /label= N_myristoylation_site
FT   Modified-site   110..113
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   117..180
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   164..167
FT                   /note= cAMP and cGMP dependent protein kinase
FT                   phosphorylation site
FT   Domain          224..233
FT                   /label= P_loop
FT                   /note= "Kinase :a domain"
FT   Modified-site   226..268
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   227..232
FT                   /label= N_myristoylation_site
FT   Domain          263..357
FT                   /label= NBS_domain
FT   Modified-site   269..272
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   290..293
FT                   /note= cAMP and cGMP dependent protein kinase
FT                   phosphorylation site
FT   Domain          290..306
FT                   /label= Walker_B_box
FT                   /note= "Kinase_2 domain"
FT   Region          325..327
FT                   /note= "RGD cell attachment sequence"
FT   Domain          344..355
FT                   /note= "Kinase 3a domain"
FT   Modified-site   347..349
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   426..428
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   433..435
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Domain          448..449
FT                   /label= Dileucine_motif
FT   Modified-site   491..496
FT                   /label= N_myristoylation_site
FT   Modified-site   522..525
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Domain          533..534
FT                   /label= Dileucine_motif
FT   Domain          559..560
FT                   /label= Dileucine_motif
FT   Modified-site   588..591
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   592..595
FT                   /note= cAMP and cGMP dependent protein kinase
FT                   phosphorylation site
```

XX
PN   WO20016005-A2.
XX
PD   23-AUG-2001.
XX
PF   16-FEB-2001; 2001WO-US40145.
XX
PR   17-FEB-2000; 2000US-0506067.
PR   01-SEP-2000; 2000US-0653901.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-514773/56.
DR   N-PSDB; AAD14323.
XX
PT   Novel polypeptide comprising NBS-1 or PYRIN-1 of the pyrin domain
PT   protein family useful in screening and detection assays and for
PT   treating, e.g., cancer, viral infections, autoimmune disease, and
PT   Alzheimer's -
XX
PS   Claim 8; Fig 4; 11pp; English.
XX
CC   The invention relates to human NBS-1 (nucleotide binding site) and
CC   PYRIN-1 protein and nucleic acids. NBS-1 and PYRIN-1 DNA, protein and
CC   their modulators are useful in the treatment of apoptotic and
CC   inflammatory disorders, cancer (leukaemia, melanoma, carcinoma); viral
CC   infections (including herpesvirus and adenovirus), autoimmune diseases
CC   (systemic lupus erythematosus (SLE), immune-mediated glomerulonephritis,
CC   arthritis); neurological disorders (Alzheimer's disease, Parkinson's
CC   disease, amyotrophic lateral sclerosis (ALS), retinitis pigmentosa,
CC   Huntington's disease); aplastic anaemia, ischaemia, meningitis, liver
CC   diseases. NBS-1 and PYRIN-1 DNA, protein and their modulators are also
CC   used for the treatment of inflammatory and immune disorders such as
CC   chronic inflammatory diseases such as Crohn's disease, insulin-dependent
CC   diabetes, organ specific autoimmunity, including multiple sclerosis,
CC   Grave's disease, allergy, asthma, HIV, tuberculosis and lepromatous
CC   leprosy. The present sequence is human PYRIN-1 protein.
XX
SQ   Sequence 1034 AA;

Query Match       24.9%;  Score 800.5;  DB 22;  Length 1034;
Best Local Similarity   28.4%;  Pred. No. 1.3e-70;
Matches 219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps 17;

```
Qy    1 MLAQPQRLFLFILDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLLVT 59
            ::                ::  :                   :       ::
Db  288 IVRKPSRILFLMDGFDELQGAFDEHIGPLCTDMQKAERGDILLSSLRKKLLPEASLLIT 347

Qy   60 TRAAAPGRLQGRJ-CSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
        ||                                  ||         ::
Db  348 TRPVALEKLQHLLDHPRHVEILGFSEAXRKEYFFKYFSDEACARAAFSLIQENEVLFTMC 407

Qy  120 FVFFVCWIVCTV-RQQ-ELGRDLSRTSKITTSVYLFITSVLSSAPVADGPRLCGDLRNL 179
        : ::::::::::    :  :::  :::::: ::: :::: :::     :: :    :
Db  408 FIPLVCWIVCTGLKQOMESGKSLAQTSKTTTAVYVFFLSSLLQPRGGSQEHGLCAHLWGL 467

Qy  180 CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
        |   |: : |||::     :        :  :    :::  :    :      :::||||
Db  468 CSLAADGIWNQKILFEESDLRNHFGZQKADV-SAFLRMNLFQKEVDCEKFYSFIHMTFQEF 526

Qy  240 LAALSYLLEDGGVPRTAAGG------VGTLLRGDAQ-PHSHLVLTTRFLFGLLSAER 289
        ::: :||||                 ||:||||     :::  ::::||||||::||
Db  527 FAAMYYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER 586

Qy  290 MRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEGEEP 349
        :    ::    :  :                         :    :
Db  587 TSYLEKKLSCKISQQIRLELLKWI-------EVKAKAKKLQ--IQPSQ------ 625

Qy  350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
               |||||||::||:||: ||:  ||:||              :::||||
Db  626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEN-LSTRMDHMVSSFCIENCHRVESLSLG 680
```

FT   Modified-site   595..597
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Domain          606..607
FT                   /label= Dileucine_motif
FT   Region          618..626
FT                   /note= "Peroxisomal targetting signal"
FT   Modified-site   624..627
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   654..657
FT                   /label= N_glycosylation_site
FT   Modified-site   656..658
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Modified-site   657..660
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   717..722
FT                   /label= N_myristoylation_site
FT   Modified-site   740..743
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Region          740..767
FT                   /label= Leucine_rich_repeat
FT   Domain          740..991
FT                   /label= LRR_domain
FT   Modified-site   750..753
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Region          769..796
FT                   /label= Leucine_rich_repeat
FT   Region          797..821
FT                   /label= Leucine_rich_repeat
FT   Domain          815..816
FT                   /label= Dileucine_motif
FT   Region          816..837
FT                   /note= "Leucine zipper pattern"
FT   Domain          823..824
FT                   /label= Dileucine_motif
FT   Region          826..849
FT                   /label= Leucine_rich_repeat
FT   Region          854..878
FT                   /label= Leucine_rich_repeat
FT   Region          888..893
FT                   /label= N_myristoylation_site
FT   Region          883..906
FT                   /label= Leucine_rich_repeat
FT   Modified-site   911..914
FT                   /label= Leucine_rich_repeat
FT   Modified-site   919..924
FT                   /label= N_glycosylation_site
FT                   /label= N_myristoylation_site
FT   Region          911..935
FT                   /label= Leucine_rich_repeat
FT   Modified-site   921..924
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Domain          929..930
FT                   /label= Dileucine_motif
FT   Domain          934..935
FT                   /label= Dileucine_motif
FT   Region          940..967
FT                   /label= Leucine_rich_repeat
FT   Modified-site   950..953
FT                   /label= N_glycosylation_site
FT   Domain          962..963
FT                   /label= Dileucine_motif
FT   Modified-site   968..970
FT                   /label= Protein_kinase_C_phosphorylation_site
FT   Region          968..991
FT                   /label= Leucine_rich_repeat
FT   Modified-site   970..973
FT                   /label= Casein_kinase_II_phosphorylation_site
FT                   /note= cAMP and cGMP dependent protein kinase
FT                   phosphorylation site
FT   Domain          997..998
FT                   /label= Dileucine_motif
FT   Modified-site   1014..1017
FT                   /label= Casein_kinase_II_phosphorylation_site
FT   Modified-site   1018..1021
FT                   /label= N_myristoylation_site

XX
CC   The invention relates to a new method for the identification of a
CC   compound that binds to a human PYRIN-1 (an apoptotic signaling
CC   molecule appearing as ABU08503) comprising contacting the polypeptide or
CC   a cell expressing the polypeptide to a test compound and determining
CC   whether PYRIN-1 binds to the test compound, or determining the effect of
CC   the test compound on the activity of PYRIN-1. In a similar manner,
CC   compounds are isolated which modulate the binding of PYRIN-1 to ASC
CC   (comprising testing the compounds against the pyrin binding domains of
CC   both PYRIN-1 and ASC (not defined)), modulate the activity of NF-kB
CC   (nuclear factor kB), and which modulate the ASC-mediated activation of
CC   NF-kB (comprising: (a) measuring the binding of a test compound to the
CC   LRR (leucine rich repeat) domain of PYRIN-1; and (b) measuring the
CC   activation of NF-kB in a cell expressing ASC and PYRIN-1 in the presence
CC   and absence of the LRR domain binding compound). The candidate
CC   modulators may be useful for treating an inflammatory disorder.
CC   Also included is a method (M8) for identifying a modulator of caspase-1
CC   activity. The identified compound can be used to modulate the
CC   polypeptide's activity (ASC and NF-kB activities in a patient). By
CC   modulating the expression or activity of the polypeptide (PYRIN-1 and
CC   ASC), a disorder associated with inappropriate apoptosis (inflammatory
CC   disorders e.g. familial cold urticaria, arthritis, inflammatory bowel
CC   disease, Crohn's disease, ulcerative colitis, rheumatoid arthritis, Lyme
CC   disease, insulin-dependent diabetes, multiple sclerosis, Grave's
CC   disease, contact dermatitis, psoriasis, graft versus host disease,
CC   asthma, chronic obstructive pulmonary disease, allergies (e.g. food
CC   allergies), conjunctivitis, helminthic infection (e.g. Leishmanias's),
CC   viral infections such as HIV infection, and bacterial infections such as
CC   tuberculosis and leprosy) can be treated. Human NBS-1 protein (nucleotide
CC   binding site protein 1) contains a pyrin domain which was used to
CC   identify genomic sequences containing the PYRIN-1 gene. The gene for
CC   pyrin-1 is located on chromosome 1q44. The present sequence represents
CC   human PYRIN-1.
XX
SQ   Sequence   1034 AA;

Query Match   24.9%;  Score 800.5;  DB 24;  Length 1034;
Best Local Similarity   28.4%;  Pred. No. 1.3e-70;
Matches 219;  Conservative 117;  Mismatches 241;  Indels 193;  Gaps 17;

Qy    1  MLAQPQRLLFILDGADELPALGGPEAAP-CTDPFEAASGARVLGGLSKALLPPTALLVT   59
            |:| ||:||:|||::  ::  :||   |||  :|: :::||:||::  :|::||:||:
Db  288  IVRKPSRILFLMDGFDELQGAPDEHIGPLCTDWQKAERGDILLSSJIRKKLLPEASLLIT  347

Qy   60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKYFYKFFRDERRAERAYRFVKENETLFALC  119
            || : :|| | |   |||  ||||::|:: ||  |||::| |||||:||::||::||: |
Db  348  TRPVALEKIQHLLDHPRHVEILGFSEAKRKEYFFKYFSDEAQARAAFSLIQENEVLFTMC  407

Qy  120  FVPFVCWIVCTLVLRQQLELGRDLSRTSKTTTSVVLLFITSVLSSAPVADGPRLQGJLRNL  179
            | | |||||||  | ||  ::|  |:|||||::| :|| ||:|::     | | : ||:|:
Db  408  FIPLVCWIVCTGLKQOMESGKSLAQTSKTTTAVYVFFLSSLLLQPRGGSQEHGLCAHLWGL  467

Qy  180  CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVYQFIDQSFQEF  239
            | | | || :::   |  | | || :  :|: : :  :    ||||| |  ||: :|||
Db  468  CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRMNLFQKEVDCEKFYSFIHMTFQEF  526

Qy  240  LAALSYLEDGGVPRTAAGG------------VGTLLRGDAQ-PHSHLVLTTRFLFGLLSAER  259
            :|| | | |: |   :       :          | | |   | | ||:||||||:: :||
Db  527  FAAMYLLEEEKEGRTNVPGSRLKLPSRDVTVLLENYGKFEKGYLIFVVRFLFGLVNQER  586

Qy  290  MRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEP  349
            :: :|:  |  :|:: :|: :  |                :     :      :
Db  587  TSYLEKKLSCKISQQIRLELLKWI-----------EVKAKAKKLQ--IQPSQ--------  625

Qy  350  NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI  409
             : || |||||| ||:||| :| : ||| :|:|| : |::|  : ||  |  | ::|:|
Db  626  --LEJFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG  680

Qy  410  SCRLVAAQEKKKKSLGKRLQ-----ASLGGGSSQQ-TTKQLPASLLHPLF----QA  455
            ||||| ||||||||||:||||      ||||||||  ||||  ||||||||    ||
Db  681  FLHNMPKEEEEEKEGRHLDMVQCVLPSSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS  740

Qy  456  MT-----------DP-----  459

---

Qy  410  SCRLVAAQEKKKKSLGKRLQ-----ASLGGGSSQQ-TTKQPASLLHPLF----CA  455
Db  681  FLHNMPKEEEEEKEGRHLDMVQCVLPSSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS  740

Qy  456  MT-----------DP-----  459
Db  741  LTELDLSDNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE  800

Qy  460  ----------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL  496
Db  801  LDLSDNALGDFGIRLLCVGLKHLLCNLKKLWLVSCCLTSACCQDLASVLSTSHSLTRLYV  860

Qy  497  LHNRLSEAGLRMLSEGLAWPCCRVQTVRV-------------  525
Db  861  GENALGSDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSA:SSVLSTNQNLTRLYLRGN  920

Qy  526  CLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM--------  561
Db  921  TLGD--KGIKLLCEGLLHPDCKLQVLELDNCNLTSHCCWDLSTLTSSQSLRKLSLGNNC  978

Qy  562  ----VTYLCAVLQHQGCGLQTLSLASVELSEQSLQEQAVKRAKPDLVI  636
Db  979  LGDLGVMMFCEVLKQQSCLLQNLGLSEVYFNYETKSALETLQEEKPELTV  1028

RESULT 23
ABU08503
ID   ABU08503 standard; protein; 1034 AA.
XX
AC   ABU08503;
XX
DT   22-MAY-2003  (first entry)
XX
DE   Huma PYRIN-1 protein.
XX
KW   Human; nucleotide binding site; pyrin domain; NBS-1; PYRIN-1;
KW   caspase-1; antiinflammatory; apoptosis; ASC; NF-kB; nuclear factor KB;
KW   LRR; leucine rich repeat; inflammatory disorder; familial cold urticaria;
KW   arthritis; inflammatory bowel disease; Crohn's disease; asthma; allergy;
KW   ulcerative colitis; rheumatoid arthritis; Lyme disease; Grave's disease;
KW   insulin-dependent diabetes; multiple sclerosis; contact dermatitis;
KW   psoriasis; graft versus host disease; food allergy; conjunctivitis;
KW   chronic obstructive pulmonary disease; food allergy;
KW   helminthic infection; Leishmaniasis; viral infection; HIV infection;
KW   bacterial infection; tuberculosis; leprosy; chromosome 1q44.
XX
OS   Homo sapiens.
XX
PN   US2002187922-A1.
XX
PD   12-DEC-2002.
XX
PF   22-APR-2002; 2002US-0.275516.
XX
PR   17-FEB-2000; 2000US-0506067.
PR   31-SEP-2000; 2000US-0653901.
PR   26-SEP-2001; 2001US-0964955.
PR   20-DEC-2001; 2001US-0027629.
XX
PA   (BERT/) BERTIN J.
PA   (MANJ/) MANJI G A.
XX
PI   Bertin J, Manji GA;
XX
DR   WPI; 2003-328763/31.
DR   N-PSDB; ABX93556.
XX
PT   Identification of compound that binds to polypeptide (for e.g. a PYRIN
PT   protein), useful for treating disorders associated with inappropriate
PT   apoptosis, for e.g. inflammatory disorder -
XX
PS   Claim 1; Fig 4; 74pp; English.

```
Db   741  LTELDLSDNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE  800
               ::        |                               :|:  ||| ::|||  |

Qy   460                    LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL  496
                           :||::|||  | | |  ::|| :||:|| | | ||||||

Db   801  LDLSDNALGDFGIRLLCVGLRHLLCNLKKLWLVSCC--SACCQDLASVLSTSHSLTRLYv  860
               |  |   ::| | | |  ||   || || : | |   ||| ||||||:|| :| ||||

Qy   497  LHNRLSEAGLRMLSEGLAWPQCRVQTVRV-----------------------------  525
              |  :||||||||||||||:|:||||:||:

Db   861  GENALGDSGVAILCEKAKPQCNLQKLGLVNSGLTSVCCSALSSVLSTNCNLTHLYLRGN  920
             |:|| :||| | | |:|  ||:||| :|||||||:||| ||| ::| |  :| :|:

Qy   526  QLPDPCRGLQYLV-GMLRQSPALTLDLSGCQLPAPM-----------------------  561
            | |  ||| ||| || |||| ||||||||||  ||

Db   921  TLGD--KGIKLLCEGLLHPDCKLCVLELDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND  978
              |||    ||||| ||| ||| |:||:|:|| |   ||:||||| :| ||||:|| ::|

Qy   562  -----VTYLCAVLCHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDLVI  606
                 ||| ||| || |:|||:||||| ::| |: |:||||:|||| |:

Db   979  LGDLGVMMPCEVLKQQSCLLQNLGLSEMYFNYETKSALETLQEEKPDLTV  1028
            |||| | ::|| | :  || |||||  |:|:  |:|| |||||:|||| |
```

RESULT 14

AAB62571

AAB62571;

ID   AAB62571  standard; Protein; 1429 AA.
XX
AC   AAB62571;
XX
DT   23-JUL-2001  (first entry)
XX
DE   Human CARD-7 polypeptide.
XX
KW   CARD-7; CARD-8; CARD-5; caspase recruitment domain; cancer; human;
KW   autoimmune disorder; antiinflammatory; immunosuppressive; antiallergic;
KW   antibacterial; antiviral; gene therapy.
XX
OS   Homo sapiens.
XX
FH   Key              Location/Qualifiers
FH
FT   Domain           329..645
FT                    /note= "nucleotide binding domain"
FT   Domain           333..341
FT                    /note= "kinase 1A subdomain"
FT   Domain           404..413
FT                    /note= "kinase 2 subdomain"
FT   Domain           454..463
FT                    /note= "kinase 3a subdomain"
FT   Domain           615..622
FT                    /note= "motif 2 domain"
FT   Domain           807..834
FT                    /note= "leucine-rich domain"
FT   Domain           836..863
FT                    /note= "leucine-rich domain"
FT   Domain           864..891
FT                    /note= "leucine-rich domain"
FT   Domain           893..920
FT                    /note= "leucine-rich domain"
FT   Domain           921..948
FT                    /note= "leucine-rich domain"
FT   Domain           950..976
FT                    /note= "leucine-rich domain"
XX
RN   WO200130813-A1.
XX
PD   03-MAY-2001.
XX
PF   27-OCT-2000; 2000WO-US29796.
XX
PR   27-OCT-1999; 99US-0428252.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
XX
PI   Bertin J;
XX
DR   WPI; 2001-343340/36.
```

DR   N-PSDB; AAF83651.
XX
PT   identifying a modulator of interaction between caspase recruitment
PT   domain (CARD)-7 and CARD-5, for treating autoimmune disorders,
PT   comprises measuring the binding of CARD-7 and CARD-5 in the presence of
PT   the compound -
XX
PS   Disclosure; Fig 1A-D; 80pp; English.
XX
CC   The invention relates to identifying compounds, that modulate interaction
CC   between caspase recruitment domain (CARD)-7 and CARD-5. The method
CC   involves measuring the binding of CARD-7 and CARD-5 in the presence of
CC   the compound (an increase in the binding of CARD-7 to CARD-5 in the
CC   presence of the compound compared to the binding in the absence of the
CC   compound indicates that the compound is a modulator of CARD-7-CARD-5
CC   interaction). Modulators of CARD-7 and CARD-8 expression or activity can
CC   be used to treat or diagnose disorders such as cancers, bacterial or
CC   viral infections, autoimmune disorders (systemic lupus erythematosus,
CC   immune-mediated glomerulonephritis or arthritis), inflammatory disorders,
CC   organ-specific autoimmunity including multiple sclerosis, Hashimoto's
CC   thyroiditis, or Grave's disease, psoriasis, graft rejection, allergies.
CC   CARD-7 and CARD-8 are useful as modulating agents in regulating a variety
CC   of cellular processes including cell growth and cell death. The present
CC   sequence represents the human CARD-7, an intracellular protein.
XX
SQ   Sequence  1429 AA;
```

Query Match             23.7%;  Score 762;  DB 22;  Length 1429;
Best Local Similarity   33.2%;  Pred. No. 1.6e-66;
Matches  220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps  22;
```

```
Qy     1  MLAQPQRLLFILDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV   58
             :::||||||||||||||  ||::|||| ||:|   |: ||||| || | ||:| |||

Db   398  LSRPERLLFILDGVDE-PGWVLQEPSSELCLHWSQPQPADALLGGSLLGKTILPEASFLI  456

Qy    59  TRAAAPGRLQGRLCSPQCAEVRGFSOXDKKKYFYKFFRDERRAERAYRFVKENETLFAL  118
             |:::|    | | |   :| |     |  :  : | | |   : |   : |  ::

Db   457  TARTTALONLIPSLEQARWVELGFSESSRKEYFVRYFTDERQAIRAFRLVKSNKELWAL  516

Qy   119  CFVPFVCWIVCTVLRQLELGRDLSRFSKTTTSVYLLFITSVSSAPVADGPRJQGDLRN  178
            ||| |:|     ||: |  : : ||  | |: |||:  |::   ||    |    |:|

Db   517  CLVPWVSWLACTCLMQQMKRKEKLTLTSKTTTLCLHYLAQALQAQPL--GPQ------LRD  570

Qy   179  LCRLAREGVLGRRAQPAEKELEQLELRGSKVQTLFLSKKELPGVLETE---VTYQFIDQS  235
            |||:|  | | :   ||||| :|  :    || ||  |:   |: ||:     || ||

Db   571  LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIST-FLKX----GILQEHPIPLSYSFIHLC  625

Qy   236  FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL-----------------T  277
            |:|| |  :       :  :  |   |: |     | ||  :                |

Db   626  FQEFFAAMSYVLED------------EKGRG------KHSNCIIDLEKTLEAYGIHGLFGAST  672

Qy   278  TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDT  337
            |||| |:|:|||    ||  : :|  :|  |:  |  |                 : : :

Db   671  TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV-----------------PSLQLL  719

Qy   338  EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC--RMDVAVLS  394
            :   :   :::  | |   :| |::||| | |  |: |:|| | :||   |: :|:||

Db   711  LQP------------HSLSLHCLYETRNKTFLTQVMAHFEEMGM----CVETDXELLVCT  755

Qy   395                 YCVRCCPAGQALRLISCR----------------LVAAQEKKKKSLGKRLQ  429
                          :| :|  | || | ||                   | | | ||| |::|

Db   756  FCIKFSRHVKKLQLIEGRQHRSTWSPTMVVLFRWVPVTDAVWQILFSVLKVTRNLKELD  814

Qy   430  ASLGGGSQGTTKQLPASLLHFLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP  489
            |||:|: :::|  ||||:||| |  | |:   |   |      |  ||| | |:|| :

Db   815  LS---GNS------LSHSAVSLCKTLRRPCLLETLRLAGCGLTAEDCKDLAFGLRANQ  865

Qy   490  ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGL-----QYLVGMLRQS  544
            :||||:|:||||||||||||||||:|:||||:||:|:|   ::|     |:| | |:|

Db   866  TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS  921

Qy   545  PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL  604
            ||| | ||||:|| |  |    | || | :|| |||:|||::| |: |:||||:|||| |
```

```
Db   922 PSLKELDLQQNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy   605 VI 606
          ::
Db   982 LI 983


RESULT 15
ABG97969
ID   ABG97969 standard; Protein; 1429 AA.
XX
AC   ABG97969;
XX
DT   07-JAN-2003 (first entry)
XX
DE   Human leucine rich repeat domain containing protein #2.
XX
KW   Leucine rich repeat; nervous system; human; neural disorder; apoptosis;
KW   renal disorder; immune disorder; arthritis; asthma; AIDS;
KW   acquired immunodeficiency syndrome; metabolic disorder;
KW   pulmonary disease; cardiovascular disease; hyperproliferative disorder;
KW   neurodegenerative disease; Alzheimer's disease; Parkinson's disease;
KW   Huntington's disease; developmental disorder; autoimmune disease;
KW   Addison's disease; haemolytic anaemia; antiphospholipid syndrome;
KW   allergic encephalomyelitis; gene therapy.
XX
OS   Homo sapiens.
XX
PN   WO200274959-A2.
XX
PD   26-SEP-2002.
XX
PF   20-DEC-2001; 2001WO-US50457.
XX
PR   03-JAN-2001; 2001US-259479P.
PR   09-JAN-2001; 2001US-260616P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Ramanathan C, Feder J, Mintier G;
XX
DR   WPI; 2002-750554/81.
XX
PT   New HLRRNS1 nucleic acids and polypeptides, useful for preventing,
PT   treating, or ameliorating e.g. renal disorder, immune, hematopoietic,
PT   metabolic, reproductive, pulmonary, cardiovascular or autoimmune
PT   diseases
XX
PS   Example 1; Page 392-397; 415pp; English.
XX
CC   The invention describes nucleic acids encoding human leucine-rich repeat
CC   containing proteins expressed in nervous system tissues, HLRRNS1. The
CC   HLRRNS1 polypeptide or the polynucleotide is useful for preventing,
CC   treating, or ameliorating a neural disorder or a disorder related to
CC   aberrant apoptosis modulation (either directly or indirectly), renal
CC   disorder, immune disorder (e.g. arthritis, asthma, acquired
CC   immunodeficiency syndrome (AIDS) or rheumatoid arthritis),
CC   haematopoietic, metabolic, reproductive, pulmonary or cardiovascular
CC   diseases, hyperproliferative disorders, neurodegenerative diseases
CC   (e.g. Alzheimer's disease, Parkinson's disease or Huntington's
CC   disease) developmental disorders, non-infectious disorders, nervous
CC   system diseases and/or disorders, and autoimmune diseases (e.g.
CC   Addison's disease, haemolytic anaemia, antiphospholipid syndrome, or
CC   allergic encephalomyelitis). The polynucleotides are also useful as
CC   chromosome markers, for chromosome identification, gene therapy, and in
CC   identifying organisms from minute biological samples. This is the amino
CC   acid sequence of a leucine-rich repeat containing protein.
XX
SQ   Sequence 1429 AA;

Query Match             23.7%;   Score 762;  DB 23;  Length 1429;
Best Local Similarity   33.2%;   Pred. No. 1.6e-66;
```

Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

```
Qy     1 MLAQPQRLLFLFLDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 58
           ::|||:|| ||| |::  |||| ::|::  :|||  || :|:|  ||  :| |||: :|
Db   398 ILSRPERLLFLFLDGVDE-PGWVLQEPSSELCLHWSQPQPADALLGSLLGKTILPEASFLI 456

Qy    59 TTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRARAYRFVKENETLFAL 118
          |:||| :| :  |:: ||| |:||:|:: :|:|| :|: | ||:|:|| : ::|:|:|
Db   457 TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFVRYFTDERQAIRAFRLLVKSNKELWAL 516

Qy   119 CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVVLLFITSVLSSAPVADGPRLQGGDLRN 178
          | || |.| .|| |:||:: :|  || :||||: || ::  || : |.||  |:   |:
Db   517 CLVPWVSWLACTCLMQQMKRKEKLTLTSKTTTTLCLHYLAQALQACPLJ-GPQ---LRD 573

Qy   179 LCRLAREGVLGRRAQFAEKELEQELRGSKVQTLFLSKKESPGVLSTE--VTYQFIDQS 235
          || |: ||: : :: ||   : | :| :| :: ::   |  :|| :     :  :  |
Db   571 LCSLAAEGIWQKKTLFSPDLRKSHGLDGAIIST-FLXM---GILQEHPIPLSYSFIHLC 625

Qy   236 FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL--------------T 277
          |||| || || |||         |  |:                          :
Db   626 FQEFFAAMSYVLED--------EKGRG-------KHSNCIIDLEKTLEAYGIHGLFGAST 670

Qy   278 TRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDT 337
          |||| |||| : :  :|  : |  |:                         | :|   :
Db   671 TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV----------------PSLQLJ 720

Qy   338 EEPEEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFC---RMDVAVLS 394
          : |             :| ||||| ::  |:::: : | |  |        :  :  | :
Db   711 LQP---------HSLESLHCLYETRNKTFLTQVMAHFEEMGM-----CVETDMELLVCT 755

Qy   395 YCVRCCPAGQALRLISCR-----------------------------LVAAQEKKKKSLGKRLQ 429
          :|. : : |: | |: ||                             :::: :::: :| ||:
Db   756 FCIKFSRHVKKLQLIEGRQHRSTWSPTMVVLFRWVPTDAYWQILFSVLVKVTRNL-KELD 814

Qy   430 ASLGGGSQGTTKQLPASLLHPLFQAVTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP 489
          :|: ::: |   :|  ||   |   : :| |: ::  :|:  ||::|  ||   ||:|:
Db   815 LS----LSHSAVKSLCKTLRRPRCLLETLRLAGCGLTAEDCKDLAFGLRANQ 865

Qy   490 ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQCVRVQLPDPQRGL----QYLVGMLRQS 544
          :|||| ||:| |::||  :|   ::   | :::  ||:|:   : |    |||| :|::|
Db   866 TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSCGLTSDCCQDLASVLSAS 921

Qy   545 PALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
          |:| :|||.  |   : | :||  |:|::|||: | |::||:|||:::||:: ::: |:|
Db   922 PSLKELDLQQNLDDVGVRLLCEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL 981

Qy   605 VI 606
          ::
Db   982 LI 983
```

Search completed: October  2, 2003, 17:48:51
Job time : 92 secs

CM protein - protein search, using sw model

Copyright (c) 1993 - 2003 Compugen Ltd.

GenCore version 5.1.6

Run on:    October  2, 2003, 17:45:33 ; Search time 107 Seconds
                                        (without alignments)
                                        1507.317 Million cell updates/sec

Title:           US-10-029-347-2
Perfect score:   3218
Sequence:        : MLAQPQRLLFILDGADFLPA.........ITHPALDGHPQPPKELISTF  625

Scoring table:   BLOSUM62
                 Gapop 10.0 , Gapext 0.5

Searched:        830525 secs, 258052694 residues

Total number of hits satisfying chosen parameters:          630525

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :       SPTREMBL_23:*
                 1: sp_archea:*
                 2: sp_bacteria:*
                 3: sp_fungi:*
                 4: sp_human:*
                 5: sp_invertebrate:*
                 6: sp_mammal:*
                 7: sp_mhc:*
                 8: sp_organelle:*
                 9: sp_phage:*
                10: sp_plant:*
                11: sp_rodent:*
                12: sp_virus:*
                13: sp_vertebrate:*
                14: sp_unclassified:*
                15: sp_rvirus:*
                16: sp_bacteriap:*
                17: sp_archeap:*

SUMMARIES

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

| Result No. | Score | % Match | Query Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 835 | 25.9 | 1033 | 11 | Q8R4B8 | Q8r4b8 mus musculu |
| 2 | 829.5 | 25.8 | 892 | 4 | Q8NEJ4 | Q8neu4 homo sapien |
| 3 | 644.5 | 20.0 | 713 | 6 | Q95LZ7 | Q95lz7 macaca fasc |
| 4 | 569.5 | 17.7 | 846 | 4 | Q8IXT0 | Q8ixt0 homo sapien |
| 5 | 473 | 14.7 | 982 | 11 | Q8BU40 | Q8bu4c mus musculu |
| 6 | 464.5 | 14.4 | 748 | 11 | Q99MW0 | Q99mw0 mus musculu |
| 7 | 429.5 | 13.3 | 825 | 11 | Q8C6M5 | Q8c6m5 mus musculu |
| 8 | 412.5 | 12.8 | 863 | 11 | Q8C6J9 | Q8c6j9 mus musculu |
| 9 | 369.5 | 11.5 | 657 | 11 | Q9EPG7 | Q9epg7 mus musculu |
| 10 | 356 | 11.1 | 673 | 11 | Q8CCN1 | Q8ccn1 mus musculu |
| 11 | 321.5 | 10.0 | 692 | 4 | Q96D51 | Q96d51 homo sapien |
| 12 | 305 | 9.5 | 1020 | 11 | Q8K3Z0 | Q8k3z0 mus musculu |
| 13 | 289 | 9.0 | 953 | 4 | Q8IWF5 | Q8iwf5 homo sapien |
| 14 | 271.5 | 8.4 | 778 | 4 | Q8NF48 | Q8nf48 homo sapien |
| 15 | 267.5 | 8.3 | 1155 | 7 | Q97PP1 | Q97pp1 mus musculu |
| 16 | 254.5 | 7.9 | 706 | 11 | Q8BUT6 | Q8but6 mus musculu |
| 17 | 254.5 | 7.9 | 953 | 11 | | Q8BHB0 | Q8bhb0 mus musculu |
| 18 | 238 | 7.4 | 1052 | 7 | Q9GJD8 | Q9gjd8 rattus norv |
| 19 | 238 | 7.4 | 1073 | 7 | Q9GJD9 | Q9gjd9 rattus norv |
| 20 | 238 | 7.4 | 1153 | 7 | Q9GJE0 | Q9gje0 rattus norv |
| 21 | 235.5 | 7.3 | 519 | 11 | Q8C249 | Q8c249 mus musculu |
| 22 | 235 | 7.3 | 660 | 4 | Q8NF06 | Q8nf06 homo sapien |
| 23 | 233 | 7.2 | 884 | 4 | Q29675 | Q29675 homo sapien |
| 24 | 233 | 7.2 | 932 | 4 | Q96KL4 | Q96kl4 homo sapien |
| 25 | 229 | 7.1 | 509 | 4 | Q9H5Z8 | Q9h5z8 homo sapien |
| 26 | 226 | 7.0 | 1056 | 4 | Q8NF42 | Q8nf42 homo sapien |
| 27 | 223 | 6.9 | 977 | 7 | Q8HW99 | Q8hw99 mus musculu |
| 28 | 213 | 6.6 | 461 | 6 | Q8H2P9 | Q8h2p9 pan troglod |
| 29 | 212 | 6.6 | 447 | 4 | Q96FD7 | Q96fd7 homo sapien |
| 30 | 212 | 6.6 | 461 | 4 | Q8IZK8 | Q8izk8 homo sapien |
| 31 | 211 | 6.6 | 461 | 4 | Q9BQ80 | Q9bq80 homo sapien |
| 32 | 211 | 6.6 | 516 | 4 | Q8NBF5 | Q8nbf5 homo sapien |
| 33 | 210.5 | 6.5 | 287 | 4 | Q9BY26 | Q9by26 homo sapien |
| 34 | 206 | 6.4 | 1097 | 4 | Q9H6Y0 | Q9h6y0 homo sapien |
| 35 | 198.5 | 6.2 | 456 | 11 | Q91VI7 | Q91vi7 mus musculu |
| 36 | 196.5 | 6.1 | 456 | 11 | Q924P4 | Q924p4 mus musculu |
| 37 | 180 | 5.6 | 390 | 11 | Q8BWZ1 | Q8bwz1 mus musculu |
| 38 | 170 | 5.3 | 1004 | 5 | Q94427 | Q94427 ciona intes |
| 39 | 169.5 | 5.3 | 733 | 4 | Q8TEE2 | Q8tee2 homo sapien |
| 40 | 162 | 5.0 | 554 | 4 | Q9H724 | Q9h724 homo sapien |
| 41 | 149 | 4.6 | 753 | 4 | Q96917 | Q96917 homo sapien |
| 42 | 146 | 4.5 | 269 | 11 | Q9D458 | Q9d458 mus musculu |
| 43 | 134 | 4.2 | 1194 | 4 | Q9H7K4 | Q9h7k4 homo sapien |
| 44 | 130 | 4.0 | 237 | 11 | Q8BT74 | Q8bt74 mus musculu |
| 45 | 130 | 4.0 | 743 | 13 | Q8JGM4 | Q8jgm4 gallus gal- |

ALIGNMENTS

RESULT 1
Q8R4B8
ID   Q8R4B8                  PRELIMINARY;       PRT;   :033 AA.
AC   Q8R4B8;
DT   01-JUN-2002 (TrEMBLrel. 21, Created)
DT   01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Mast cell maturation inducible protein 1.
GN   MMIG1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/cJ;
RA   Kikuchi-Yanoshita R., Koga K., Taketomi Y., Sugiki T., Saito T.,
RA   Ishii S., Hisada M., Suzuki-Nishimura T., Uchida M.K., Moon T.-C.,
RA   Chang H.-W., Sawada M., Inagaki N., Nagai H., Murakami M., Kudo I.;
RT   "Identification of inducible genes during in vitro maturation of mouse
RT   bone marrow-derived mast cells to connective tissue-type mast cells.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF486632; AAL90874.1; -.
DR   InterPro; IPR001092; HLH_basic.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   InterPro; IPR007091; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   SMART; SM00368; LRR_RI; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS00038; HLH_1; 1.
DR   PROSITE; PS50503; LRR_RI; 3.
DR   PROSITE; PS50837; NACHT; 1.
SQ   SEQUENCE 1033 AA;  118274 MW;  5924690966B12117 CRC64;

Query Match             25.9%;  Score 835;  DB 11;  Length 1033;

DR    InterPro; IPR007039; LRR_RNinh.
DR    InterPro; IPR007111; NACHT_NTPase.
DR    InterPro; IPR004020; PAAD_DAPIN_dom.
DR    Pfam; PF02758; PAAD_DAPIN; 1.
DR    PROSITE; PS50824; DAPIN; 1.
DR    PROSITE; PS50503; LRR_RI; 1.
DR    PROSITE; PS50837; NACHT; 1.
SQ    SEQUENCE   892 AA;  101733 MW;  AE703D8DF341C2AC CRC64;

Query Match             25.8%;
Best Local Similarity   33.1%;
Matches  215;  Conservative 109;  Mismatches 241;  Indels 24;  Gaps 14;

Score 829.5; DB 4; Length 892;
Pred. No. 2.3e-57;

Qy      1  MLAQPQRLLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT  59
           :|: :|||||| |||||    | |  | | |||||||||||||| ||||||||||||||
Db    282  LIRVPERLLFIIDGFDE-KPSFHDPQGPWCLCWEEKRPTELLNSLIRKLLPELSLLIT  341

Qy     60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC  119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    342  TRPTALEKLRHLLEHPRHVEILGFSEAERKEYFYKYFHNAEQAGQVFNYVRDNEPLFTMC  401

Qy    120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNL  179
           ||||||||||||| ||| |||   |   || ||||||| |||  ||   ||  ||||
Db    402  FVPLVCWVVCTCLQQQLEGGGLLRQTSRTTAVVMLYLLSLMQPKPGA--PRLQPPPNQR  459

Qy    178  NLCRLAREGVLGRRAQPAEKELEQLELRGSKVQT-FLSKKELPGVLETEVTYQFIDQSFQ  237
           |  |  ||| |||| |  |||  ||  |   |  |||||| |||||  |  | ||| ||
Db    460  GLCSLAADGLWNQKILFEEQDLRKHGLDGEDV-SAFLMNIFQKDINCERYYSFIHLSFQ  518

Qy    238  EFLAALSYLLEDGGVPRTAAGGVG------TLLRGDAQPHSHIVLTRFLFGLLSAERM  290
           ||  | |    ||||  |  |          |   |   |   | |  |||||||
Db    519  EFFAAMYYILDEG------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLFGLLNETR  572

Qy    291  RDIERHFGCMVSERVKQEALRWVQGGQGCPGVAPEVTEGAKGLEDTEPEEEGEGEPN  350
           :|:|  :|| | | |  |  |  |  |  |||  | |             |    |
Db    573  SHLEKSLCWKVSPHIKMDLLQWIQSKAQ---------------------SDGSTLQQGS--  610

Qy    351  YPLELLYCLYETQEDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAGQALRLIS  410
           :|  | || | ||  | |  | |  |  :   | |   |  ||  |||   | ||  |
Db    611  --LEFFSCLYEIQEEEFIQQALSHFQVIVVSNIA-SKMEHMVSSFCLKRCRSAQVLHLYG  667

Qy    411  CRLVAAQEKKKKSLGKRLQASLGGGS--------------SQGTKQLPASL-  448
           || :  |||||||| | | |  | |              ||| ||| |:|
Db    668  ATYSADGE-------DRARCSAGAHTLLVQLRPERTVLLDAYSEHLAAALCTNPNLIELS  720

Qy    449  LHPLFQAVTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLH  498
           |  | |  |||||| |  |||||| | |    |  || | | || | ||
Db    721  LYRNALGSRGVKLLCQGLRHPNCKLQNLRLKRCRISSSACEDLSAALIANKNJTRMDLSG  780

Qy    499  NRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTLDLSGCQL  557
           ||| | |  | | ||  | || || || |        | |  ||  | | |||| ||
Db    781  NGVGFPGWMLLCEGLRHPQRCRLQMQLRKCQLESGACQEMASVLGTNPHLVELDLTGNAL  840

Qy    558  PAPMVTYLCAVLQHQGCGLQTL-SLASVELSEQSLQELCAVKRAKPDLVI  606
           |  :   || | |  |  |    ||    ||  || ||| | |  |||:|
Db    841  EDLGLRLLCQGLRHPVCRLRTLWVLFGMDLNKMTHSRLAALRVTKPYLDI  890

RESULT 3
Q95LZ7
ID   Q95LZ7           PRELIMINARY;       PRT;   713 AA.
AC   Q95LZ7;
DT   01-DEC-2001 (TrEMBLrel. 19, Created)
DT   01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical 80.3 kDa protein.
OS   Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC   Cercopithecinae; Macaca.
OX   NCBI_TaxID=9541;
RN   [1]
RP   SEQUENCE FROM N.A.

Best Local Similarity   29.3%;   Pred. No. 1e-57;
Matches  226;  Conservative 112;  Mismatches 239;  Indels 194;  Gaps 15;

Qy      1  MLAQPQRLLFILDGADELP-ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT  59
           :|: :|||||| |||||    | |  | | ||||||||||||| ||||||||||||||
Db    286  ILRKFSRILFLMDGFDELQGAFDEHIGEVCTDWQKAVRGDILLSSLIRKLLPKASLLIT  345

Qy     60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC  119
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    346  TRPVALEKLCHLLDHPRHVEILGFSEAKREYFYKFYFSNELQAREAFRLQENEVLFTKC  405

Qy    120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTSVLLFITSVLSSAPVADGPRLQGDLRNL  179
           ||| ||||||||| ||| |||   |   || |||||||  ||   |     ||  ||||
Db    406  FIPLVCWIVCTGLKQQMETGKSLAQTSKTTAVVFFLSLLQSRGGIEEHLFSCYLQSL  465

Qy    180  CRLAREGVLGRRAQPAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF  239
           | |  ||| |||  |||  |  |   |    |  ||||||  |||  |  | |||| ||
Db    466  CSLAADGIWNQKILFEECDLRKHGLQKTDV-SAFLRMNVPQKEVDCERFYSFSKMTFQEF  524

Qy    240  LAALSYLLE---DGGVPRTAAGGVGTLLRGDAQ--------PHSHLVLTRFLFGLLSA  287
           |  | |||    ||||  |  |  |  || |         || |||  | ||||||
Db    525  FAAMYLLEEEAEGETVRKGPGGCSDLLNRDVKVLENYGKFEKGYLIFVVRFLFGLVNQ  584

Qy    288  ERMRDIERHFGCMVSERVKQEALRWVQGGQGCPGVAPEVTEGAKGLEDTEPEEEEGE  347
           |||  |  || | | ||| |  | |  |  |  ||||||| |  |||||| |  |
Db    585  ERTSYLEKKLSCKISQQVRLELLKWI------------EVKAKAKKLQ--WQPSQ---  625

Qy    348  EPNYPELLYCLYETQEDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAGQALR  407
           :  | |||| | ||  | |  | | |  | :   | |   |  ||  |||   | ||
Db    626  --LELFYCLYEMQEEDFVQSAMDHFFKIEIN--LSTRXDHVVSSFCIKNCHRVKTLS  678

Qy    408  LSCRLVAAQEKKKKSLGKRLQ---------ASLGGG--SSQGTK  442
           | |  |  |||||||| | |          ||       | |
Db    679  LGFFHNSPKEEEERRGGRPLDQVQCVFPDTHVACSSRLVNCCLTSSFCRGLFSSLSTR  738

Qy    443  QL---------PASLLHPLFQAMTDPLCHLSSLITLSHCKLPDAVCRDLSEALRAAPA  490
           |           | ||| |    |||||| |  |||||| | |    |  ||
Db    739  SLTELDLSDNTLGDPG--MRVLCEALQHPGCNIQRLWLSRCGLSHQCCFDISSVLSSSGK  796

Qy    491  LTELGLLHNRLSEAGLRMLSEGLAWPQCR  519
           | | ||| ||| | |  | | ||  | ||
Db    797  LVEILDLSDNALGDFGIRLLCVG0KHLLCHLQKLWVSCCLTSACCQDLAIVLSSNHSJTR  856

Qy    520  ---------VQTVRVGLPDPQCRGLQYL-  537
           ||| | |  |
Db    857  LYIGENALGSGVQVLCEXMKDPQCNLQKIGLVNSGLTS-CCSA-TSVLKTNQNFTHJYL  916

Qy    538  ------VGMLRQSPALTTLDLSGCQLPAP-  560
           || |   | || |
Db    917  RSNALGDTGIRLLCEGLLHPDCKLQMLEJDNCSLTHSHSCKNLSTILTHHNHSLRKLNLGNN  976

Qy    561  -----MVTYLCAVLQHQGCGLQTCEVLKQGGCLLQSLQLGEXYLNRETKRALEALQEEKPELTI 1027
Db    977  DLGDLCVVTLCEVLKQGGCLLQSLQLGEXYLNRETKRALEALQEEKPELTI 1027

RESULT 2
Q8NEU4
ID   Q8NEU4           PRELIMINARY;       PRT;   892 AA.
AC   Q8NEU4;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Monarch-1 splice form IV.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RA   Williams K.L., Linhoff M.W., Ting J.P.Y.;
RL   Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AY116207; AAM75145.1; -.

```
RC   TISSUE=Testis;
RA   Hashimoto K., Osada N., Hida M., Kusuda J., Tanuma R., Hirai K.,
RA   Terao K., Sugano S.;
RT   "Isolation of novel full-length cDNA clones from macaque testis cDNA
RT   libraries.";
RL   Submitted (SEP-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AB071044; BAB64437.1; -.
DR   InterPro; IPR007091; LRR_RNint.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   PROSITE; PS50503; LRR_RI_2.
DR   PROSITE; PS50837; NACHT_1.
KW   Hypothetical protein.
SQ   SEQUENCE  713 AA;  80258 MW;  6F214C9B773F54DC CRC64;
```

```
Query Match            20.0%;  Score 644.5;  DB 6;  Length 713;
Best Local Similarity  31.2%;  Pred. No. 9.8e-43;
Matches 205;  Conservative  95;  Mismatches 248;  Indels 109;  Gaps  18;
```

```
Qy    1 MLAQPQRLLFILDGADELPALGGPEAA---PCTDPFEAASGARV.GGLLSKALLPTALL   56
         |||:  :: |:||.||: ::|| ::|| |  |||  :||:| ::::|||
Db   16 ILAQAQKILFVIGFDE---IGAPPGAIQDICGDWEQQKPVPVLLGSLLKKMIRKAII   72

Qy   57 LVTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLF   116
         ::: |||: :|| :|||||| ::||::::   |: ||:|:||||||||||||||||||
Db   73 LVTTRPRALRDLRILAEQPIYIRVEGFLEEDRRAYFLRHFGDEDQAVRAFELMRSNAALF   132

Qy  117 ALCFVPFVCWIVCTVLRGRDLSRTSKTTSVILFITSVLSSAPVADGPRLQGDL   175
         |:|  | |||||  |:|  ::|:||:|||:||::|||||||||||||:
Db  133 QLGSAPAVCWIVCTTKLQMEKGEDPAPTCLTSTGLFRFLCS----QPPGAQIRGAL   187

Qy  177 RNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQF.DQSF   236
         |:||||| |:|||  :|||||||| || || |||  ::||||||||||:||| :||
Db  188 RALSLLAAQSLWAQMSVLHGEDELESAGVQESDLR-LFLDGDILRQDGVAKGCYSFIHLSF   246

Qy  237 QEFLAALSY:LE------DGGVPRTAAGGVGTLL-RGDAQPHSHLVLTTRFLFGLLSAE   288
         ||| :: | :    |||||||||||||| |:| || :||| :|||
Db  247 QQFLTALFYALEKEEEDRDGHA--WDIGDVQKLLSREERLKNPDLIQAGRFLFGLANEK   304

Qy  289 RMDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEEEGE   348
         | |::| |  || |||||   |:||::|  |   |
Db  305 RVKELEATFGWRMSPEIKQELGRCDVSRKNGHFTAA--------DLR-----   343

Qy  349 PNYPLELLYCLYETQEDAFVRQALCRPPELALQRVRFCRMDVAVLSYCVRCC---PAGQ   404
         : : | || ::|| ||||||| |    || ||: ||
Db  344 ----ELLCCLYESQEDELVKEVMAQFKEISL--HLNAVDIAPSSPCFKHCQNIFPAD-   394

Qy  405 ALRLISCR-LVAAQEKKKKKSLGKRLQAS-----------LGGGSSQGTT   441
         | |:: |:  || ||||||||||:||||
Db  395 -----AGRNLCLALRGHKTVTHLTLQGTDQKDMLPALCEVLRHPECNLRYLGLVSCSATT   449

Qy  442 KQ------LPAS----------LLHP----LFQAWTDPLCHLSSLTLSHCKLPDAVC   478
         ::       |: ::    :|| :||||||||||||||||||:|||
Db  450 CQWAQDLS.ALEANRS:XCVNLSDNELLDEGAKLLYTTLRHPKCFLQRLSLENCHLTEANC   509

Qy  479 RDLSEALRAAPALTEIGLLHNRLSEAGLRMLSEGLAWPQCRVQCTVRVQLPD-PQRGLQYL   537
         |:|:|||||||::|| :||:||||||||||||||||||:|||:|| |||||   ||||||
Db  510 KDLAAVLVVSRELTHLCLAKNSLKDTGVKFLCEGLSYPECKLQALVLWNCDITSDGCCSL   569

Qy  538 VGMLRQSPALTTLDLSGGCQLPAPMVTYLCAVLQHQGGGLQTLSLASVELSEQSLQE.   594
         |::|:||||||||| ||| ||| |:|||||||||||||||::|||:||||:|
Db  570 AKLLQEKSSLSCLDLGLNHIGVTGVKVLCEALSKPLCNLRCLWLWGGSIPPFSCEDL   626
```

```
RESULT 4
Q81XT0
ID   Q81XT0;      PRELIMINARY;      PRT;   846 AA.
AC   Q81XT0;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to NALP2 protein.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
```

```
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RA   Strausberg R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC039269; AAH39269.1; -.
SQ   SEQUENCE  846 AA;  96368 MW;  46BE24555DE39F7 CRC64;
```

```
Query Match            17.7%;  Score 569.5;  DB 4;  Length 846;
Best Local Similarity  30.3%;  Pred. No. 1.2e-36;
Matches 192;  Conservative  92;  Mismatches 233;  Indels 117;  Gaps  19;
```

```
Qy    1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARV.LGGLLSKALLPTALLLVT   59
         |||:  :: |:||.||: ::|| ::||  |||  :||:| ::::|||
Db  276 ILAQARKILFVIDGFDELGAAPGALIEDICGDWEKKKPVPVLLGSLLNRVM.PKAALLVT   335

Qy   60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC   119
         ||||: :|| :|||||| ::||::::   |: ||:|:||||||||||||||||||
Db  336 TRPPRALRDLRILAEEPIYIRVEGFLEEDRRAYFLRHFGDEDQAMRAFELMRSNAALFQLG   395

Qy  120 FVPFVCWIVCTVLRQCLE.GRDLLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNL   179
         | | |||||  |:|  :|::|:||:|||:|||:||::|||||||||||||:
Db  396 SAPAVCWIVCTTLKLQMEKGEDPVPTCLTRTGLFRFLCS----RFPQGAQLRGALRTL   450

Qy  180 CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF   239
         ||||| |:|||  :||||||||||| || || |||  ::|||||||||:|||:||
Db  451 SLLAAQGLWAQTSVLHREDLERLGVQESDLR-LFLDGDILRQDRVSKGCYSFIH.SFQQF   509

Qy  240 LAALSYLLE-------DGGVPRTAAGGVGTLLRG-DAQPHSHLVLTTRFLFGLLSAERMR   291
         | :: |     |||||||||||||| |:| || :||| :|||
Db  510 LTALFYTLBKEEEEDRDGHT--WDIGDVQKLLSGVERLRNPDLIQAGYYSFGLANEKRAK   567

Qy  292 DIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEEGEEPNY   351
         | |::| |  || |||||   |:||::|  |   |
Db  568 ELEATFGCRMSPDIKQELLRC---DISCKGGHSTVTD---------LQ---   603

Qy  352 PLELLYCLYETQEDAFVRQALCRPPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISC   411
         : : | || ::|| ||||||| |    || ||: ||
Db  604 --ELLGCLYESQEEELVKEVMAQFKEIS---HLNAVDVVPPSFCVKHC--RNLQKMSL   655

Qy  412 RLVAAQEKKKKKSLGKRLQASLGGGS---------SQGTTKQLP-   445
         | |:: |:  || |||||||||:||||
Db  656 QVI------KENLPENVTASESDAEVERSQDDQHMLPFWTDLCSIFGSNKDLMGLAINDS   709

Qy  446 ---ASLLHLPLFQAWTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPAL.ELGLLHNRLS   502
         :|||  |: :||||||||||||:|||
Db  710 FLSASLVRILCEQIASDTCHLQRVVFKNISPADA-HRNLCLALRGHKTVTYLTLQGNDQD   768

Qy  503 EAGLRMLSEGLAWPQCRVQCTVRVQLPDPQRGLQYLVGM-------------LRQSPA   546
         |||:|||||||||||:|||:|| |||||  ||||||||
Db  769 DM-FPALCEVLRHPEC---------------NLRYL-GLVSCSA.TQQWADLS.ALEVNQS   812

Qy  547 LTTLDLSGGCQLPAPMVTYLCAVLQHQGGCGLQTLS   580
         |||||| ||| ||| |:|||||||||||||||::
Db  813 LTCVNLSDNELLDEGAKLLYTTLRHPKCFLQRLS   846
```

```
RESULT 5
Q8BU40
ID   Q8BU40;      PRELIMINARY;      PRT;   982 AA.
AC   Q8BU40;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to PAN2 protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
```

RC    STRAIN=C57BL/6J; TISSUE=Ovary;
RX    MEDLINE=22354683; PubMed=12466852;
RA    The FANTOM Consortium;
RA    the RIKEN Genome Exploration Research Group Phase I & II Team;
RT    "Analysis of the mouse transcriptome based on functional annotation of
RT    60,770 full-length cDNAs.";
RL    Nature 420:563-573 (2002).
DR    EMBL; AK087843; BAC40024.1; -.
SQ    SEQUENCE 982 AA; 112600 MW; 6E2DBCFEA2053585 CRC64;

Query Match             14.7%;  Score 473; DB 11; Length 982;
Best Local Similarity   23.5%;  Pred. No. 7.1e-29;
Matches 174;  Conservative 116; Mismatches 254; Indels 198; Gaps 20;

Qy      1 MLAQPQRLLFILDGADELPA-LGPEAAPCTDPFEAASGARV.GGLLSKALLPTALLVT 59
Db    218 ILSKPEKLLFIIDSLEGMECDLFKWESELCDNCTEKQPVNVLLSSLLRRKMLPESSLLIS 277

Qy     60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
Db    278 ATPESFEKMENRIEYTHVKI-KGLKERNIKMSFHRLFQDRNRAHEAFSLVRENEQLFTVC 337

Qy    120 FVPFVCWIVCTV;RQCLE.GRDLSRTSKTT'SVYL------LFITSVLLSSAPVADGPRLCG 174
Db    338 QVPVLCWMVATCLKEEIEKGRDPVSICRCTTSLYTTHIFNLFIPQNA4SPSKKSQDQLQG 397

Qy    175 DLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQ 234
Db    398 ----LCSLAAEGMWTDTFVFGEEALRRNGIMDSDIPIL-LDIGMLINIRESEKSYIFLHP 452

Qy    235 SFCEFLAALSYLLEDG-GVPRTAGGVGTLLRGDAQP-HSHLVLTTRFLFGLLSAERMRD 292
Db    453 SVQEVCAA;FYLLKSHVDHPSQEVKSIEKLMFAFLKKVKVQWIFFGSFIFGLLHESEQKK 512

Qy    293 IERHFGCMVSERVKQEALRWVQGQGQGPGVAPEVTEGAKGLEDTEPPEEEEGEEPNYP 352
Db    513 .EAFFGHQ.SQEIKRQ.YQCLE-----------TISGNEELQEQ:DG------- 548

Qy    353 LELLYC:YETQEDAFVRQAL.CRFPEPELALQRVRFCR-------MDVAVLSYCVRCCPAGQALRL 408
Db    549 MKLFYCLFEMDDDTFLVEAVNC------MEQINFVAKDYSDVIVAAHCLKHC---FTLKK 599

Qy    409 ISC-----------QEKKKKSLG- 425
Db    600 :SFSTQNVLSGAQEHSYMERLLITAWNHICSVF;FISKDIQELRMKDTNLSESAFSVLITNNL 659

Qy    426 ------------KRLQASLGGGS- 436
Db    660 KYHNYTLNVLVANNVFFVCEKYLFFELIQNCNLQHLNLSLTIL5HSDVKLLCDVLSQAEC 719

Qy    437 --------SQGTTKQLPASL----:HPFLFQAMTDPLCHLSS 465
Db    720 NIEELVVAACSISSDDCKVFASVLISNKTLKHNLASNTLDKGlASLCKSLCHPDCILEH 779

Qy    466 --------RGLQYLVGM:RCSPALTTLDLSGCQLPAPXVTY:CAV:QHC 572
Db    780 :VLANCSLNEKCWDYlSEVLRRNKTLSHLDISSNDLKDEGLKV:------CRALT-- 828

Qy    573 GCGLQTLSLASVELSEQSLQEL 594
Db    829 -LPDSVLKSLSLRHCLITISGCQDLAEV:RNNQN:VSLQVSNNKLEDTGVKLLCDAIKHP 887

Db    888 NCHLEDLGLEACELTGACCEDL 909

RESULT 6
Q99YW0
ID   Q99YW0                 PRELIMINARY;   PRT;   748 AA.
AC   Q99YW0;
DT   01-JUN-2001 (TrEMBLrel. 17, Created)
DT   01-JUN-2001 (TrEMBLrel. 17, Last sequence update)

DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Ribonuclease/angiogenin inhibitor 2.
GN   RNH2.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Testis;
RX   MEDLINE=21175748; PubMed=11279525;
RA   Wang P.J., McCarrey J.R., Yang F., Page D.C.;
RT   "An abundance of X-linked genes expressed in spermatogonia.";
RL   Nat. Genet. 27:422-426(2001).
DR   EMBL; AF285581; AAK13960.1; -.
DR   HSSP; P10775; 2BNH.
DR   MGD; MGI:1890518; Rnh2.
DR   InterPro; IPR007091; LRR_RNin.
DR   PROSITE; PS50501; LRR_RI 1.
SQ   SEQUENCE 748 AA; 84946 MW; 674A385E011DE8EC CRC64;

Query Match             14.4%;  Score 464.5; DB 11; Length 748;
Best Local Similarity   22.9%;  Pred. No. 2.4e-28;
Matches 158;  Conservative 119; Mismatches 238; Indels 175; Gaps 17;

Qy     41 VLGGLLSKALLPTALLVTTRAAAPGRLQGRLCSPQCAEVR---GFSDKDKKKYFYKFFR 97
Db     25 LLSSLLRRKMLPKSSFLISATPETFBKMEGRV--ECTNVKIVTGFNESNIKMYFRSLFQ 81

Qy     98 DERRAERAYRFVKENETLFALCFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFI 157
Db     82 DKTKTQEIFSLVKENQQLFTVCQVPVLCWMVATCLKKEIEKGRDLVSVCRRTTSLYTTHI 141

Qy    158 TSV-LSSAPVADGPRLQGDLRNLCRLAREGV.GRRAQFAEKELEQLELRGSKVQTLFLSK 216
Db    142 FNLFIPQSAQYPSKESQAQLQSLCSLAAEGMWTDTFVFGEEALRRNGIMDSDIPTL--LDV 203

Qy    217 KELPGVLETEVTYQFIDQSFDEFLAALSYLLED-----GGVPRTAAGGVGTLLRGDAQPH 271
Db    201 RILEKSKKSEKSYIFLHPSIQVCAAIFYLLKKSHMDHPSQDVKSIEALIFTFLK---KVK 257

Qy    272 SH:VLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGA 331
Db    258 VQWIFFGSFIFGLLHESEQKKLEAFFGHQLSQEIKRQLYQCLE---- 300

Qy    332 KGLEDTEPEEEEGEEPNYPLELLYCLCYETQSDAFVRQAL-CRFPEE.ALQRVRFCR-- 387
Db    301 -TISGNEELQEQVDG------MKLFYCLFEMDDEAFLAQAMNC-- 347

Qy    389 MDVAVLSYCVRCCPAGQALR---------------LISCRLVAAQE 418
Db    348 SDVIVAAHCLQHCSTLKKLSLSTQNV:SEGQEHSYTEKLLMCWHHMCSVLISSKDIYIZQ 407

Qy    419 KKKKSLGKRLQASLGGGS------------- 436
Db    408 VKSNTNLNE--TASLVIYSHLMYPSCTLKALVVNNV'FLCDNRLFFELIQNCCLQHLDNL 465

Qy    437 ---SQGTTKQL------------PASLLHP 451
Db    466 TF.SHGDVKLLCDVLSQEECNIEKLMVAACNLSPDDCKVFASVLISSKMLKHLNLSSNNL 525

Qy    452 ------:FQAMTDPLCHLSSLT.SHCKLPDAVCRD.SEALRAAPALTELGLLHNRLSEAG 505
Db    526 DKGISSLSKALCHPDCVLKNLVLVNCSLSEQCWDYLSEVLRRNKTLRHLDISSNDLKDEG 585

Qy    506 LRMLSEGLAWPQCRVQTVRVQ-LPDPQRGLQYLVGMLRQSPALTTLDLSGCCLPAPMVTY 564
Db    586 LKVLCRALSLPDSVLKSLSVRYCLITTSGCDLAEVLRKNQNLRNLQVSNNKIEDAGVKL 645

Qy    565 LCAVLQHQGCGLQTLSLASVELSEQSLQEL 594
Db    646 LCDAIKHPNCHLENIGLEACALTGACCEDL 675

```
RESULT 7
Q8C6M5
ID   Q8C6M5      PRELIMINARY;      PRT;   825 AA.
AC   Q8C6M5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Weakly similar to PAN1.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Ovary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK054264; BAC35770.1; -.
SQ   SEQUENCE 825 AA;  94021 MW;  11D71DEABEAFCA31 CRC64;
```

Query Match   13.3%;    Score 429.5;  DB 11;  Length 825;
Best Local Similarity  25.4%;  Pred. No. 1.7e-25;
Matches 163;  Conservative 105;  Mismatches 238;  Indels 135;  Gaps 21;

```
RESULT 8
Q8C6J9
ID   Q8C6J9      PRELIMINARY;      PRT;   863 AA.
AC   Q8C6J9;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Hypothetical RNI-like structure containing protein.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=C57BL/6J; TISSUE=Ovary;
RX   MEDLINE=22354683; PubMed=12466851;
RA   The FANTOM Consortium,
RA   the RIKEN Genome Exploration Research Group Phase I & II Team;
RT   "Analysis of the mouse transcriptome based on functional annotation of
RT   60,770 full-length cDNAs.";
RL   Nature 420:563-573 (2002).
DR   EMBL; AK054426; BAC35775.1; -.
KW   Hypothetical protein.
SQ   SEQUENCE 863 AA;  99955 MW;  D3A46C782DA544CF CRC64;
```

Query Match   12.8%;    Score 412.5;  DB 11;  Length 863;
Best Local Similarity  22.6%;  Pred. No. 4e-24;
Matches 165;  Conservative 100;  Mismatches 245;  Indels 221;  Gaps 19;

Qy    560  PM-------------------------------VTYLCAV:QHQGCGLQTLSLASVELSECS-    590
Db    785  ACCKDLASTFTRYKCLRRINLAKNSLGFSGLFVLCKAMKDCTCTLYELKLRMADFDSDSSQ    844

Qy    591  ---LQELCAVK 598
Db    845  EFLLSEMERNK 855

RESULT 9
Q9EPG7
ID    Q9EPG7;                    PRELIMINARY;              PRT;     657 AA.
AC    Q9EPG7;
DT    01-MAR-2001 (TrEMBLrel. 16, Created)
DT    01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT    01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE    RNI-like protein.
OS    Mus musculus (Mouse).
OC    Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC    Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX    NCBI_TaxID=10090;
RN    [1]
RP    SEQUENCE FROM N.A.
RC    STRAIN=129/SvJ;
RX    MEDLINE=21310002; PubMed=11416212;
RA    Lane R.P., Cutforth T., Young J., Atharasiou M., Friedman C.,
RA    Rowen L., Evans G., Axel R., Hood L., Trask B.J.;
RT    "Genomic analysis of orthologous mouse and human olfactory receptor
RT    loci.";
RL    Proc. Natl. Acad. Sci. U.S.A. 98:7390-7395(2001).
DR    EMBL; AF32:233; AAG45188.1; -.
DR    InterPro; IPR007091; LRR_RNinh.
DR    InterPro; IPR007711; NACHT_NTPase.
DR    PROSITE; PS50503; LRR_RI; 1.
DR    PROSITE; PS50837; NACHT; 1.
SQ    SEQUENCE    657 AA;    75410 MW;    F4BED9E4BA19AAF3 CRC64;

Query Match           11.5%;    Score 369.5;    DB 11;    Length 657;
Best Local Similarity   25.8%;    Pred. No. 7.4e-21;
Matches   126;   Conservative   83;   Mismatches   194;   Indels   85;   Gaps   11;

Qy    135  QLELGRDLSRTSKTTSVYLLF-TSVLSSAPVADGPRLQGD--LRNLCRLAREGVLGRRA    192
                                :::  :: :  :::     :  : :  ::::
Db    163  QVEKGGRVEETCKTSTALFTYYICSLFPRIPVG-CVTLPNETLLRSLCKAAVEGIWTVKG    221

Qy    193  CFAEKELECLELRGSKVQTLFLSKKELPGVLETEVTYQF:DQSFQEFLAALSYL----LE    248
                                ::::   ::::::       ::::
Db    222  VLYQQNLRKGELTREDI-LLFLDAXVLQQCTEYENCYMFHLHVQEFFAALFYLLRENLE    280

Qy    249  DGGVPRTAAGGVGTLLRGDAQPHSH:V:TRFLFGLLSAERMEDIERHFGCM:SERVKQE    308
                                  :::    ::::  :::  : :    :::: :
Db    281  EQYPSEPFENLY:LLESNH:HDPHLEQMKCF:FGLLNKDRVRQLEETFNLTISMEVREE    340

Qy    309  ALRWVQGQGGCPGVAPEVTEGAKGLEDTEPEEEGEEPNYPLELLYCLYETQEDAFV    368
                                 ::::  :: :::::::: :::: ::::::
Db    341  :LACLEGL-----------------EKDESSLSQLRFQDLLHCIYETQDQSF:    376

Qy    369  RQALCRFPELALQRVRFCRMD----VAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSL    424
                                :::::::    :::::      :::::
Db    377  TQALMYFQKIIV-----RVDEEPQLRIYSFCLRHCHTLKTMRLTA------RADLKRM    423

Qy    425  GKRLQASLGGGSSQGTTKQLPASLLHPLFQAM:DP:CHLSS::LSHCKLPDAVCRDLSEA    484
                                                        :::
Db    424  LDTAEMCLEGAAVQ-----VIH-----YWQDLFSV    448

Qy    485  LRAAPAJTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQS    544
Db    449  LHTNESJIEMDLYESRLDES:MKILNEELSHPKCKLJQKLIFRSVDFLNGCQDFT-FLASN    507

Qy    545  PALTTJDLSGCQLPAPMVTY:CAVLQHQGCGLQT:SLASVELSEQSLQELQAVKRAKPDL    604
Db    508  KKVTHLDLKETDLGVNGLKT:CEALKCKGCKLRVLRLASCDLNVARCQKLSNALQTNRSL    567

DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Similar to hypothetical protein FLJ21478.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Bone;
RA   Strausberg R.;
RL   Submitted (AUG-2001) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BCC-3199; AAH13199.2; -.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   SMART; SM00368; LRR_RI; 6.
DR   PROSITE; PS50503; LRR_RI; 1.
KW   Hypothetical protein.
SQ   SEQUENCE   692 AA;  76206 MW;  AD9D10E7F4A7BDDE CRC64;

Query Match
Best Local Similarity   10.0%;  Score 321.5;  DB 4;  Length 692;
Matches  174;  Conservative  85;  Mismatches 255;  Indels 211;  Gaps 27;

```
Qy    1 MLAQPQRLLFILDGADELP---ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLL  57
          |||| :| : | ::|:|||    ||  |  :|| :|  |  |::::|||||||:|| | :
Db    1 MAAAGSHLLFV.HGLEHLNLCFRLAG--TGLCSDPEEPQEPAAI?VNLLRKYMLFQASLL  58

Qy   58 VTTRAAAPGRLQGRLCS?QCAEVRGFSDKDKKKYFYKFFRDE---------PRAE     103
          ||| ||: ||:||  | | :  || ||::||     || ||           |||
Db   59 VTTRPSAIGRIPSKYVG-RYGEICGFSDTNLQKLYFQLRLNQPYCGYAVGGSVSA?PAQ  117

Qy  104 RAY------RFVKENETLFALCFVPFVCWIVC?VLRQQLELGRDLSRTSKT?SVYLFI   157
          |         |::  : :|| |  |||| ::  : |    |  | |  | | | :: |:
Db  118 RDHLVOMLSRNLEGHHQIAAACFLPSYCWLVCAT---LHFLHAPTPAGQT.TSIY?SF?  173

Qy  158 T-------SVLSSAPVADGERLQGDLRNLCRLAREGVLGRRAQFA----------EXE   198
                  | |||:    | | ||  | || : | :    |   ||             |
Db  174 RLNFSGETLDSTDPSNLSLMAYAARTMGKLAYEGVSSRKT?VFSEEEVCGCLEAGIRTEEE 233

Qy  199 LEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFLAALSYLLEDGGVPRTAAG  258
          : || :  :: : | |    | | ||  |  | || | |||| | | |||||  ||| |
Db  234 FQLLHIFRRDALRFFLAPCVEPG--RAGTFVTVPAMQEYLAALYIVL--GLRKT?LQ     287

Qy  259 GVG---TLLRGDAQPHSHLVL-------TTRFLFGLLSAERMRDIERHFGCMVS---    302
          ||    | | | | : | :::        ||:| |:|:    |      | | |
Db  288 KVGKEVAELVGRVGEDVSLVLGIMAKLLPLRALPLLFNLIKV----VPRVFGRMVGKSR  342

Qy  303 ERVKQE-ALRKVGQGQGCPGVAPEVTEGAKGLE----DTEEPEEEEGE---------  347
          |  ||    |:||   |  | || :    | ::     | |||:: |
Db  343 EAVAQAMVLEMFREEDYNEDVLDQMGASILGVEGPRRHPDEPPEDEVFELFPMFMVGGLL 402

Qy  348 --------------------------------ENYPLELLYCLYETQEDAFV        368
                                          | | || |  || |:|  ||
Db  403 SAHNRAVLAQLGCPIKNLDALENAQAIKKKLGKLGRQV.LPFSELLDHLFFHYEFQNQRFS 462

Qy  369 RQALCRFPELALQRVRF----CRMDVAVLSYCVRCCPAGQ---ALRLISCRLVAAQEK-  419
          : |:  | | |::|        | |: |:  || |  |     :|: : |:|   |:
Db  463 AEVLSSLRQLNLAGVRMTPVKCTVVAAVLG------SGRHALDEVNLASCQLDPAGLRT   515

Qy  420 -----KKKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKL  473
               :::||| :|  ||    | |   :|| : |  |       | :||: ||||| |
Db  516 LLPVFLRARKLGLQLN-SLG---PEACKDLRDLLLH-------DQCQITTLRLSSNPL    562

Qy  474 PDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG  533
          | |   ||  || |  ||| ||||| ||  || :||:|||          ||     |
Db  563 TEAGVAVLMEGLAGNTSVTHLSLLHTGLGDEGLELLAAQL----------DRNRQ      607

Qy  534 LQY?VGMLRQSPALTT?DLSG?QLPAPMVTYLCAV?CHQGCG-LQTLSLASVELSECSLQ  592
          ||:  | | :    :|  ||                  | |   | |:||:   |   |:
Db  608 LQEL------------------------------NVAYNGAGDTAALALARAAREHPSLE  637

Qy  593 ELQAV 597
          :|||| 
Db  638 LLQGV 642
```

---

RESULT 12
Q8K3Z0
Q8K3Z0
ID   Q8K3Z0       PRELIMINARY;       PRT;  1020 AA.
AC   Q8K3Z0;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   CARD15.
GN   CARD15.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090; .
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/c;
RA   Iwanaga Y., Davey M.P., Martin T.M., Planck S.R., DePriest M.L.,
RA   Baugh M.M., Suing C., Rosenbaum J.T.;
RT   "Cloning, sequencing and expression analysis of the murine Nod2/Card-5
RT   gene.";
RL   Submitted (JUN-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AF520774; AAM76073.1; -.
DR   MGD; MGI:2429397; Card15.
DR   InterPro; IPR001315; CARD.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR001130; TatD_DNase.
DR   PROSITE; PS50209; CARD; 2.
DR   PROSITE; PS50503; LRR_RI; 1.
DR   PROSITE; PS50837; NACHT; 1.
DR   PROSITE; PS01137; TATD_1;1.
SQ   SEQUENCE  1020 AA; 113561 MW;  25549905EC770FB3 CRC64;

Query Match
Best Local Similarity    9.5%;  Score 305;  DB 12;  Length 1020;
Matches  157;  Conservative 111;  Mismatches 277;  Indels 124;  Gaps 28;

```
Qy    1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT?  60
          :|    :|| | || | :|  || :   |:  :    |   |:  |||: |   |||| :
Db  347 LLDHDRVLLTFDGLDEFKFRTDRERHCS-PIDPTSVQTLFNLLQGN-LKNACKVLTS    405

Qy   61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENET?FALCF 120
          |   | || ::    | | ::  |  :| | |  ::|||:  | |||:|||||| |||| 
Db  406 RPDAVSALLRKFVRTE-LQLKGFSEEGIQLYLRKHHREPGVADRLIQIQATSALHGLCH  464

Qy  121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVL------SSAPVADGPRL-  172
          :||  :::  :|  :| | ||  ||    | :: |::  |:       |:   : |||
Db  465 LPVFSKMVSRCHRELLLQNRGFPTTS---TDMYLLILQHFLLHASPPDSSPLGLGPGLLQ 521

Qy  173 --QGDLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFL--SKKELPGVLETEVT 228
            |   :|| : | : :|     |:: :| || :|:: : :||    :  :|   :||:
Db  522 SRLSTLLHLGHLALRGLAMSCYVFSAQQLQAAQVDSDDISLGFLVRAQSSVPG---SKAP 578

Qy  229 YQFIDQSFQEFLAA-------------------LSYLLEDGGV                 259
          ::|:  :| :|:||                   || || |
Db  579 LEFLHITFQCFFAAFYLAVSADTSVASLKHLFSCGRLGSSLLGRLLPNLCIQGSRVKKGS 638

Qy  260 VGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQG 319
          : ::|   |:| | ::: :::| |||| ::  |    :  ::|||    | :      |
Db  639 EAALLQ-KAEPH-NLQITAAFLAGLLS-QHRDLLA--ACQVSERV---LLQRQARARS   689

Qy  320 CPGVAPEVTEGAKGLEDTEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELA   379
          |     :   |  :|      |||       | :::::::|::| |  :|   |  :
Db  690 C--LAHSLREHFHSIPPAVPGETKSMHAMPGF-IWLIRSLYEMQEEQLAQEAVRRL-DIG 745

Qy  380 LQRVRFCRM---DVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLG-----KRLQASL 432
          |:|  |||    | | |:::  :  | | |: :| || |:||         :|:|:|
Db  746 HLKLTFCRVGPAECAALAFVLQHLQRPVALQL-------DYNSVGDVGVEQLRPCL      794

Qy  433 G----------GGSSQGTTKQLPASLLHPLFQAMTDPLC-HLSSLTLSHCKLPDAVCRDL 481
```

Db   795 GVCTALYLRDNNISDRGART-VECAL------RCEQLQKLALFNNKLTRACACSM 843

Qy   482 SEALRAAPALTELGLLHNRLSEAGLRMLSEGLA----------WPQCRVQTVRVQLPDPQ 531

Db   844 AKLLAHKQNFLSLSRVGNNHITAAGAEVLAQGLKSNTSLKFLGFWGN---------SVGD- 892

Qy   532 RG-QYLVGMLRQSPALTTLDLSGCQLPAPMVTY-CAVLQHQGCGLQTLSLASVELSEQSL 591

Db   894 KGTQLAAEVVADHQNLKWLSLVGNNI-GSMGAEALALMLEKNKSLEELCLEENHLCDGSV 952

Qy   592 QEL-QAVKR 599

Db   953 YSLAEGLKR 961


RESULT 13
Q8IWF5
ID   Q8IWF5          PRELIMINARY;   PRT;    953 AA.
AC   Q8IWF5;
DT   01-MAR-2003 (TrEMBLrel. 23, Created)
DT   01-MAR-2003 (TrEMBLrel. 23, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   Caspase recruitment domain family, member 4.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Lymph;
RA   Strausberg R.;
RL   Submitted (NOV-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; BC040339; AAH40339.1; -.
SQ   SEQUENCE 953 AA; 107671 MW; 0A9DF167BE87E21A CRC64;

Query Match           9.0%;  Score 289;  DB 4;  Length 953;
Best Local Similarity 24.2%;  Pred. No. 3.2e-14; -
Matches 167;  Conservative 91;  Mismatches 276;  Indels 156;  Gaps 27;

Qy     1 MLAQPCRLLFILDGADELFA---LGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALL 57
Db   272 LLRFPHVALPTFDGLDELHSDLDLSRVPDSSC--PWEPAHPLVLLANLLSGKLLKGASKL 329

Qy    58 VTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFPRDERRAERAYRFVKENETLFA 117
Db   330 LTART---GIEVPRQFLRKKVLLRGFSPSHLRAYARRVFPERAL-CDRLLSQLFANPNLCS 386

Qy   118 LCFVPFVCWIVCTV---LRQLELGRDLSRTSKTTTSVVLLFITSV--------L 161
Db   387 LCSVPLFCMIIFRCEQHFRAAFEGSPQLPDCTMLTDVFLL-VTEVHLNRMQPSSLVQPN 445

Qy   162 SSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELEQLERGSKVQTLFLSKKELP--RALPE 220
Db   446 THSPVETLHAGRDTLCSLGQVAHRGMEKSLFVFTCEEVQASGLGERDVQLGFL-RALPE 503

Qy   221 -GVLETEVTYQFIDQSFCELAALSYLLED--------GGVPRTAA----- 257
Db   504 LGPGPGDQQSYEFFHLTLQAFFTAFFJVLDDRVGTQELLRFFQEKMPPAGAATTSCYPPFL 563

Qy   258 ---GGVGTLLRGDAQPHSHLVLTTRF-FGLLSAERMRDIERHFGCMVSERVKGEALRW 312
Db   564 PFQCLQGSGPAREDLFKNKDHFQFTNLFLCGLLSKAKQK-LLRHLVPAAALRRGRKAL-W 621

Qy   313 ---VQGQGQGCPGVAFEVTEGAKGLEDTEEPEEEESGEEPNYFLELLYCLYETQGD- 365
Db   622 AGHFSSLRGVLYLKSLPRVQVESFNQVQAM--------PTF-IWMLRCIYETQSQK 666

Qy   366 --AFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRJ-SCRLVAAQEKKKKS 423
Db   667 VGQLAARGICA-NYLKLTYCNACSADCSALSFVLHHFPXRLALDLDNNNL--------- 715

Qy   424 LGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSE 483

Db   716 ---------------NDYGVRE------LQPCFSRLT------VLRLSVNQITDGGVKVLSE 750

Qy   484 ALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDQ--PQRGLQYL- 537

Db   751 EJTKYKIVTYLGLYNNQITDVGARYVTKIL--DECKGLA-HLKLGKNKITSEGGKYLALA 807

Qy   538 --------------VGM-------------LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQ 572

Db   808 VKNSKSISEVGMGNQVGDEGAKAFAEALRNHPSLTTLSLASNGISTEGGKSLARALQ-Q 866

Qy   573 GCGLQTLSLASVELSE--QSLQELQAVKR 599

Db   867 NTSLEILWLTQNELNDSVAESLAEMLKVNQ 896


RESULT 14
Q8NF48
ID   Q8NF48          PRELIMINARY;   PRT;    778 AA.
AC   Q8NF48;
DT   01-OCT-2002 (TrEMBLrel. 22, Created)
DT   01-OCT-2002 (TrEMBLrel. 22, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   FLJ00348 protein (Fragment).
GN   FLJ00348.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Spleen;
RA   Jikuya H., Takano J., Kikuno R., Nagase T., Ohara O.;
RT   "The nucleotide sequence of a long cDNA clone isolated from human
RT   spleen.";
RL   Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
DR   EMBL; AK090431; BAC03412.1; -.
DR   InterPro; IPR001395; Aldo/ket_red.
DR   InterPro; IPR000566; Lipocln_CytFABP.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   SMART; SM00368; LRR_RI; 2.
DR   PROSITE; PS00063; ALDOKETO_REDUCTASE_3; 1.
DR   PROSITE; PS00213; LIPOCALIN; 1.
DR   PROSITE; PS50503; LRR_RI; 1.
DR   PROSITE; PS50837; NACHT; 1.
FT   NON_TER 1
SQ   SEQUENCE 778 AA; 85297 MW; A1EAAC43FF32836C CRC64;

Query Match           8.4%;  Score 271.5;  DB 4;  Length 778;
Best Local Similarity 24.2%;  Pred. No. 6e-13; -
Matches 142;  Conservative 87;  Mismatches 208;  Indels 149;  Gaps 26;

Qy     2 LAQPQRLLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
Db   225 VAVPARALLILDGLDECRTPLDFSNTVACTDPKKEIPVDHLITNIIRGNLFPPEVSIWITS 284

Qy    61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDER---------RAERAYRFVK 110
Db   285 RFSASGQIPGGLVD-RMTEIRGFNEEEIKVCLEQMFPEDQALLGWM-SQVCADRA----- 338

Qy   111 ENETLFALCFVPFVCWIVCTVLRQQLELGRDLSRTS------KTTSVLLFITSV- 161
Db   339 ----LYLMCTVPAFCRLTGMA-G--HLWR-SRTGPQDAELWPPRTLCELYSWYFRMAL 389

Qy   162 S------SAP---VADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELE---QLELR 205
Db   390 SGEGQEKGKASPRIEQVAHGGRKM--VGTLGRLAFHGLLKKKYVFYEQDMKAFGVDLALL 447

Qy   206 GSKVQTLFLSKKELPGVLETEVTYQFIDQSFDEFLAALSY--------LLEDGGV-P 253
Db   448 QGAPCSCFLQREE---TLASSVAYCFTHLSLQEFVAAAYYYGASRRAIFDLFTESGVSWP 504

```
Qy    254 RTAAGGVGTLLRGDAQ------PHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQE 308
              |||   :||||||||||   ||| |:  ::::
Db    505 RL---GPLTHFRSAAQRAMQAEDGRLDVFLRFLSGLLS-PRVNALLA--GSLLAQGEHQA 558

Qy    309 ALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEEGEEPNYPLELLYCLYETCEDAFV 368
              |                    ----------INVLHCLHELQ--- 592
Db    559 YRTQVAELLQGC--LRPDAAVCARA------------------INVLHCLHELQ--- 592

Qy    369 RQALCRPELALQRVRFCRM---DVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSL 424
              :                    :||  :|| |:  :||
Db    593 HTELARSVEEAMESGALARLIGPAHRAALAYLLQVSDA----------CAQE------ 634

Qy    425 GKRLQASLGGGSSQGTTKQLFASLLH----------PLFQAMTDPL----CHLSSLT 467
              :|                    :::|:
Db    635 ----ANLSLSLSQGVLQSLLPCLLYCRKLRLDTNQFQDPVMELLGSVLSGKLCRIQKIS 689

Qy    468 LSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRM:SEGL 513
              ||||||||||||||||||||||||||||||||||||
Db    690 LAENQISNKGAKALARSLLVNRSLTSLDLRGNSIGPGQGAKALADAL 735
```

```
Db    571 TARPR--GRLAQSLSKADAIFEVPSFSTKQAKTYMRHYFENSGTAGNQDKALGLLEQPL 628
              |||                    :::
Qy    115 LFALCFVPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFTCSVLLSSAPVADGPRLQG 174
Db    629 LCSYSHSPVVCRAVCQLSKALLEQGTE-AQLPCTLTGLYVSLLGPAAQNSP-------PG 680
              |||                    :|| |:  :::
Qy    175 DLRNLCRLAREGVLGRRAQFAEKEJEQELRGSKVQTLFLSKKELPGVLE-TEVTVQFID 233
Db    681 ALVELAKLAWE-LGRRHQ---STLQETRFSSVEVKTWAVTQGLMQQTLETTEAQLAFSS 735
              :||                    :|| :|
Qy    234 QSFQEFLAALSYL---LEDGGVPRTAAGGVGTLLRGDAQPHSH-LVLTTRFLFGLLSAE 289
Db    736 FLLQCFLGAVWLAQCNEIKDKELPQY-LA-----LTPKKKRPYDNWLEGVPRFLAGLVFQP 790
              :|| ||                    :|  :|:
Qy    289 RMRDIERHFGCMVSERV-----KQEAL-RWVGQGQGCPGVAPEVTEGAKGLEDTEPE 341
Db    791 RAHCL---GAJVEPAVAAVADRKQKVLTRYLKRLKLG------------TLRAGRL---- 831
              :|| :||                    :|::||:
Qy    342 EEEGEEPNYPLELLYCLYETQEDAFVRQALCRFP-ELALQRVRFCRMDVAVLSYCVRCC 400
              :|||                    :::|:
Db    832 --------LSLLHCAHETQQPGIWEHVAHQLPGHLSFLGTRLTPPDVYVLGRALETA 880
```

```
Qy    402 PAGQALRL---------ISC-----------RLVAAQEK- 419
              :||                    :|:
Db    881 SQDFSLDLRQTGVEPSGLGNLVGLSCVTSFRASLSDTYALWESLQQQGEAQLLQAAEEKF 940

Qy    420 ----KKKS----LGKRLQASLGGGSSQGTTKQLPA---------SLLRPLFQAMTDP- 459
              ||||                    :::|:::|
Db    941 TIEPFKAKSPKDVEDLDRLVQTQRLRNPSEDAAKDLPAIRDLKKLEFALGPLGPQAFPT 1000

Qy    460 ------------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRML 509
              :||                    :::||
Db   1001 LAKILPAFSSLQHLDLDSLSENKIGDKGVSKLSATFPQLKALETLNLSQNN:TDVGACKL 1060

Qy    510 SSGLAWPQCRVQTVRVQLPDP---QRGLQYLVGMLRQSPALLTLDLSGCQLPAPMVTYLC 566
              :||                    :|| |:
Db   1061 AEAL--PALAKSLLRLSLYNNICCDKGAKSLAQVLPDMVSLRVMDVQFNKFTAAGAQQLA 1118

Qy    567 AVLQHQGC-GLQTLSLASVELS---EQSLQELQA 596
              :||                    |:|:
Db   1119 SSLQK--CPQVETLAWWTPTIPFGVQEHLQQLDA 1150
```

Search completed: October 2, 2003, 17:51:18
Job time : 113 secs

---

```
RESULT 15
Q9TPP1
ID   Q9TPP1        PRELIMINARY;         PRT;   1155 AA.
AC   Q9TPP1; O78109; O78036;
DT   01-MAY-2000 (TrEMBLrel. 13, Created)
DT   01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT   01-MAR-2003 (TrEMBLrel. 23, Last annotation update)
DE   MHC class II transactivator CIITA.
GN   C2TA OR CIITA.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A., AND ALTERNATIVE SPLICING.
RC   STRAIN=BALB/C;
RX   MEDLINE=97327562; PubMed=9184229;
RA   Muhlethaler-Mottet A., Otten L.A., Steimle V., Mach B.;
RT   "Expression of MHC class II molecules in different cellular and
RT   functional compartments is controlled by differential usage of
RT   multiple promoters of the transactivator CIITA.";
RL   EMBO J. 16:2851-2860(1997).
CC   -!- ALTERNATIVE PRODUCTS: 3 ISOFORMS; I (SHOWN HERE), III AND IV; ARE
CC       PRODUCED BY ALTERNATIVE SPLICING.
DR   EMBL; AF000007; AAB92365.1; -.
DR   EMBL; AF100709; AAF06838.1; -.
DR   EMBL; AF100710; AAF06839.1; -.
DR   EMBL; AF042158; AAC34366.1; -.
DR   EMBL; AF042159; AAC34367.1; -.
DR   MGD; MGI:108445; C2ta.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   PROSITE; PS50503; LRR_RI; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Activator; Alternative splicing;
KW   Transcription regulation.
FT   VARSPLIC       1     77       MISSING (IN ISOFORM III).
FT   VARSPLIC       1    101       MISSING (IN ISOFORM IV).
FT   VARSPLIC      78     94       SLQAPTVERGTSYRDHG -> MRCLVPGPSSGSYLPELQ
FT                                 (IN ISOFORM III)
SQ   SEQUENCE   1155 AA;  127527 MW;  F3FF05DEBFB0CE7: CRC64.
```

Query Match              8.3%;  Score 267.5;  DB 7;  Length 1155;
Best Local Similarity   24.6%;  Pred. No. 2.1e-12;
Matches   17;  Conservative   88;  Mismatches 282;  Indels 153;  Gaps 30;

```
Qy    : MLAQPQRLLFILDGADELPALCGPEAAPC--TDPFEAASGARVLGLLSKALLPTALLLV 58
              ||||                    :|: |:
Db    512 IVRQPDRVLLLDAFEELEAQIGLLHGPCGSLSP-EPCSLRGLLAGIFQRKLLRGCTLL: 570
              :||                    :::
Qy     59 TTRAAAPGRLQGRLCSPQCA-EVRGFSDKDKKYFYKFFRDERRA---ERAYRFVKENET 114
```

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - protein search using sw model

Run on: October 2, 2003, 17:46:03 ; Search time 29 Seconds
(without alignments)
911.872 Million cell updates/sec

Title: US-10-029-347-2
Perfect score: 3218
Sequence: 1 MLAQPQRLLFILDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 328717 seqs, 42310658 residues

Total number of hits satisfying chosen parameters: 328717

Minimum DB seq length: 0
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/iaa/PCTUS COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfilesl.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | DB | Length | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 304.5 | 9.5 | 2 | 1130 | US-08-519-547A-6 | Sequence 6, Appli |
| 2 | 289 | 9.0 | 4 | 953 | US-09-099-041A-8 | Sequence 8, Appli |
| 3 | 289 | 9.0 | 4 | 953 | US-09-245-281-8 | Sequence 8, Appli |
| 4 | 289 | 9.0 | 4 | 953 | US-09-340-359B-8 | Sequence 8, Appli |
| 5 | 289 | 9.0 | 4 | 953 | US-09-340-620A-8 | Sequence 8, Appli |
| 6 | 251.5 | 7.8 | 4 | 953 | US-09-245-281-43 | Sequence 43, App |
| 7 | 251.5 | 7.8 | 4 | 953 | US-09-207-359B-43 | Sequence 43, App |
| 8 | 251.5 | 7.8 | 4 | 953 | US-09-340-620A-43 | Sequence 43, App |
| 9 | 224.5 | 7.0 | 4 | 966 | US-09-207-359B-47 | Sequence 47, Appl |
| 10 | 216 | 6.7 | 2 | 456 | US-08-910-731-2 | Sequence 2, Appl |
| 11 | 216 | 6.7 | 2 | 456 | US-08-795-395-2 | Sequence 2, Appl |
| 12 | 214 | 6.7 | 2 | 456 | US-08-910-731-8 | Sequence 8, Appli |
| 13 | 212 | 6.6 | 2 | 461 | US-08-910-731-6 | Sequence 6, Appl |
| 14 | 209.5 | 6.5 | 2 | 456 | US-08-910-731-4 | Sequence 4, Appl |
| 15 | 209.5 | 6.5 | 2 | 456 | US-08-795-395-4 | Sequence 4, Appli |
| 16 | 151 | 4.7 | 4 | 490 | US-09-099-041A-26 | Sequence 26, Appl |
| 17 | 151 | 4.7 | 4 | 490 | US-09-245-281-26 | Sequence 26, Appl |
| 18 | 151 | 4.7 | 4 | 490 | US-09-207-359B-26 | Sequence 26, Appl |
| 19 | 151 | 4.7 | 4 | 490 | US-09-340-620A-26 | Sequence 26, Appl |
| 20 | 137.5 | 4.3 | 4 | 483 | US-09-904-615-154 | Sequence 154, App |
| 21 | 126 | 3.9 | 4 | 200 | US-09-099-041A-11 | Sequence 11, Appl |
| 22 | 126 | 3.9 | 4 | 200 | US-09-245-281-11 | Sequence 11, Appl |
| 23 | 126 | 3.9 | 4 | 200 | US-09-207-359B-11 | Sequence 11, Appl |
| 24 | 126 | 3.9 | 4 | 200 | US-09-340-620A-11 | Sequence 11, Appl |
| 25 | 113.5 | 3.5 | 2 | 759 | US-08-637-759B-89 | Sequence 89, App |
| 26 | 113.5 | 3.5 | 3 | 759 | US-08-871-355A-89 | Sequence 89, Appl |
| 27 | 113.5 | 3.5 | 4 | 759 | US-09-201-945-89 | Sequence 89, Appl |
| 28 | 104.5 | 3.2 | 4 | 633 | US-09-252-991A-21659 | Sequence 21659, A |
| 29 | 104.5 | 3.2 | 4 | 1129 | US-09-252-991A-29927 | Sequence 29927, A |
| 30 | 103 | 3.2 | 3 | 740 | US-09-022-983-5 | Sequence 5, Appli |
| 31 | 103 | 3.2 | 4 | 921 | US-09-252-991A-20327 | Sequence 20327, A |
| 32 | 103 | 3.2 | 4 | 1411 | US-09-252-991A-23628 | Sequence 23628, A |
| 33 | 102.5 | 3.2 | 4 | 617 | US-09-252-991A-22318 | Sequence 22318, A |
| 34 | 102 | 3.1 | 4 | 905 | US-09-340-620A-52 | Sequence 52, Appl |
| 35 | 99 | 3.1 | 1 | 2509 | US-08-469-005A-10 | Sequence 10, Appl |
| 36 | 98 | 3.0 | 3 | 803 | US-09-154-750A-85 | Sequence 85, Appl |
| 37 | 98 | 3.0 | 3 | 2511 | US-09-261-907-2 | Sequence 2, Appli |
| 38 | 97.5 | 3.0 | 1 | 2756 | US-08-375-709-11 | Sequence 11, Appl |
| 39 | 97.5 | 3.0 | 3 | 2756 | US-08-752-929-11 | Sequence 11, Appli |
| 40 | 97.5 | 3.0 | 3 | 2756 | US-09-090-793-7 | Sequence 7, Appli |
| 41 | 97.5 | 3.0 | 4 | 2756 | US-09-231-899-7 | Sequence 7, Appli |
| 42 | 97 | 3.0 | 3 | 912 | US-08-943-768-2 | Sequence 2, Appli |
| 43 | 97 | 3.0 | 3 | 912 | US-09-865-960-2 | Sequence 2, Appli |
| 44 | 96.5 | 3.0 | 4 | 977 | US-09-302-812-2 | Sequence 2, Appli |
| 45 | 96.5 | 3.0 | 4 | 977 | US-09-511-477-2 | Sequence 2, Appli |

ALIGNMENTS

RESULT 1
US-08-519-547A-6
; Sequence 6, Application US/08519547A
; Patent No. 5994082
; GENERAL INFORMATION:
; APPLICANT:
; TITLE OF INVENTION: Proteins Essential for the Expression of
; TITLE OF INVENTION: Vertebrate MHC Class II Genes, DNA Sequences Encoding Same
; TITLE OF INVENTION: and Pharmaceutical Compositions
; NUMBER OF SEQUENCES: 6
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: FISH & NEAVE
; STREET: 1251 AVENUE OF THE AMERICAS
; CITY: NEW YORK
; STATE: NEW YORK
; COUNTRY: U.S.A.
; ZIP: 10020-1104
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: MS-DOS
; SOFTWARE: WordPerfect 6.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/519,547A
; FILING DATE: 25-AUG-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: EP94113378.7
; FILING DATE: 26-AUG-1994
; ATTORNEY/AGENT INFORMATION:
; NAME: HALEY, JAMES F.
; REGISTRATION NUMBER: 27,794
; REFERENCE/DOCKET NUMBER: VOS-11
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-596-9000
; TELEFAX: 212-596-9090
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1130 amino acids
; TYPE: amino acid
; STRANDEDNESS: double
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; HYPOTHETICAL: NO
; ANTI-SENSE: NO
US-08-519-547A-6

Query Match 9.5%; Score 304.5; DB 2; Length 1130;
Best Local Similarity 25.5%; Pred. No. 1.9e-23;
Matches 174; Conservative 80; Mismatches 262; Indels 167; Gaps 29;

Matches 167; Conservative 91; Mismatches 276; Indels 156; Gaps 27;

```
Qy     1 MLAQPQRLLFILDGADELPA---LGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLL 57
           :||  ||| |||| || ||   ||  ||
Db   272 LLRFPHVALFTFDGLDELHSDLDLSRVPDSSC--PWEPAHPLVLANLLSGKLLKGASKL 329

Qy    58 VTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFA 117
         :    ||| |
Db   330 LTART--GIEVPRQFLRKKVLLRGFSPSHLRAYARRMFFPERALQCRLLSQLEANPNLCS 386

Qy   118 LCFVPFVCWIVCTV--LRQQLEGRDLSRTSKTTTSVYLLPFITSV--------L 161
         || ||| ||| |      ||  ||||
Db   387 LCGVPLFCWIIFRCFQHFRAAFEGSPQLPDCTNLTTDVFLL-VTEVHLNRMQFSSLVQRN 445

Qy   162 SSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELP- 220
              |  ||   ||  |    |  ||  |
Db   446 TRSPVETLHAGRDTLCSLGQVAHRGMEKSLFVFTQEEVQASGLQERDMQLGFL--RALPE 503

Qy   221 -GVLETEVTYQFIDQSFQEFLAALSYLLED----------GVPRTAA----- 257
                 |   |    |   |
Db   504 LGPGGDQQSYEFFHLTLQAFFTAFFLVLDDRVGCQELLRFFQEWMPPAGAATTSCYPPPL 563

Qy   259 ----GGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCXVSERVKQEALRW 312
              ||  |   |  |  |  | |  ||
Db   564 PFQCLQGSGPAREDLFKNKDHFQFTNLFLCGLLSKAKQK-LLRHLVPAAALRKRKAL-W 621

Qy   313 VQGGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLYETQED- 365
                                                 |   |  ||| | ||
Db   622 AHLFSSLRGYLKSLPRVQVESFNQVQAM---------------PTF-IKWLRCIYETQSQK 666

Qy   366 -AFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKS 423
            | |                   |        |   |  |||
Db   667 VGQLAARGICA-NYIKLTYCNACSADCCSALSFVLHHFPKRLALDLDNNNL--------- 715

Qy   424 LGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSE 483
                                                |||| ||| ||
Db   716 --------NDYGVRE------LQPCFSRLT----------VLRLSVNQITDGGVKVLSE 750

Qy   484 ALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD--PQRGLQYL- 537
              |    |  |  |      | |  |                  |    ||| |
Db   751 ELTYKIVTYLGLYNNQITDVGARYVTKIL--DECKGLT-HLKLGKNKITSEGKYLALA 807

Qy   538 ---------VGM----------------LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQ 572
                     |||                |
Db   808 VKNSKSISEVGMWGNCVGDEGAKAFABALRNHPSLTTLSLASNGISTEGGKSLARALQ-Q 866

Qy   573 GCGLQTLSLASVELSE--QSLQELQAVKR 599
            ||  | |
Db   867 NTSLEILWLTQNELNDEVAESLAEMLKVNQ 896
```

RESULT 2
US-09-099-041A-8
; Sequence 8, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-076001
; CURRENT APPLICATION NUMBER: US/09/099,041A
; CURRENT FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-099-041A-8

Query Match 9.0%; Score 289; DB 4; Length 953;
Best Local Similarity 24.2%; Pred. No. 7.1e-22;

```
Qy     1 MLAQPQRLLFILDGADELPALGGPEAAPC-TDPFEAASGARVLGGLLSKALLPTALLLVT 59
           :||  ||| |||| || ||   ||  ||
Db   487 ILKRPDRVLILDAFEELEAQDGFLHSTCGPAPAFPCSLRGLLAGLFQKKLLRGCTLLLT 546

Qy    60 TRAAAPGRLQGRLCSPQCCA-EVRGFSDKDKKKYFYKFFRDERRAE--RAYRFVKENETL 115
Db   547 ARPR--GRLVQSLSKADALFELSGFSMEQAQAYVXRYFESSGMTEHCQRALTLRDRPLL 604

Qy   116 FALCFVPFVCW-VCTV.RQCLELGRDLSRTSKTTTSVYL-LFITSVLSSAPVADGPRLQG 174
Db   605 LSRSHSPTLCRAVCQLSEAILELGED-AKLPSTLGLYVGLLGRAALDSSP--------G 655

Qy   175 DLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSK---KELPGVLETEVTY-C 230
Db   656 ALAELAKLAWE--LGRRHQ---STLGEDQFPSADVRTWAXAKGLVQHPPRRAAESSLAFPS 710

Qy   231 FIDQSFQEFL-AALSYLLECGGVPRTAAGGVGTLLRGDAQPHSH-LVLTTRFLFGLLSAE 288
Db   711 FLLQCFLGALWLALSGEIKEKELPQYLA------LTPRKKRPYDNWLEGVPRFLAGLIFQP 765

Qy   289 RMRDIERHFGCMVSERV--KQEAL.RWVQGGQGCPGVAPEVTEGAKGLEDTEEPEEEEE 345
Db   766 PARCLGALLGPSAAAVDRKQKVLARYLKRLDPG-------TLRARQL------------- 806

Qy   346 GEEPNYPLELLYCLYETQEDAFVRQALCRFP-ELA.QRVRFCRMDVAVLSYCVRC----- 399
Db   807 ------LELLCAHEAEAGIWQHVVQELPGRLSFLGTRLCPPDAHVLGALEAAGQCP 859

Qy   400 -----CPAG-QALRLISC--------------------RLVAAQEK---- 419
Db   860 SLCLRSTGICPSGLGSLVGLSCVTRFRAAALSDTVALWESLRQHGETKLLQAAEEKFTIEP 919

Qy   420 -KKKS-----LGKRLQASLGGGSSQGTTKQLPA---------SLLHPLFQAMTDP----- 459
Db   920 FKAKSLKDVEDLGKLVQTQRTRSSSEETAGELPAVRDLKKLEFALGPVSGPQAFPKLVRI 979

Qy   460 -------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEG- 513
Db   980 LTAFSSIQHLDLDALSENKIGDEGVSQLSATFPQLKSLETLNLSQNNITDGAYKLAEAL 1039

Qy   514 AWPQCRVQTVRVQ---------LPD---------------PQRGLQYLVGMLR 542
Db  1040 --PSLAASLRLSLYNNCICDVGAESLARVLPDMVSLRVMDVQYNKFTAAGQQLAASLR 1097

Qy   543 QSPALTTLDLSGCQLPAPMVTYL 565
Db  1038 RCPHVETLAMWTPTIPFSVQEHL 1120
```

RESULT 3
US-09-245-281-8
; Sequence 8, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/019,942
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 953
; TYPE: PRT

## Left column (RESULT 4)

```
; ORGANISM: Homo sapiens
US-09-245-281-8

Query Match      9.0%; Score 289; DB 4; Length 953;
Best Local Similarity   24.2%; Pred. No. 7.1e-22;
Matches  167; Conservative  91; Mismatches 276; Indels 156; Gaps 27;

Qy    1 MLAQPQRLLFILDGADELPA---LGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLL  57
          ||||| || || | || : | |||  | |  |: ::| |   || || | | |||| |
Db  272 LLRFPHVALFTFDGLDELHSDLDLSRVPDSSC--PWEPAHPLVLLANLLSGKLLKGASKL 329

Qy   58 VTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFA 117
          | || |  |: ||::|| | |  |  :|  | :  || || | |||
Db  330 LTART--GIEVPRQFLRKKVLLRGFSPSHLRAYARRMFPERALQDRLLSQLEANPNLCS  386

Qy  118 LCFVPFVCWIVCTV--LRQQLELGRDLSRTSKTTSVLLFITSV-------L 161
          || ||| ||| :  | |  || |  |  |||
Db  387 LCSVPLFCWIIFRCFQHFRAAFEGSPQLPDCTMTLTDVFLL--VTEVHLNRMQFSSLVQRN 445

Qy  162 SSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKEL-EQLE.RGSKVQTLFLSKKELP- 220
          ||| | | || | | |  | :| |  |  || :  ||:  |
Db  446 TRSPVETLHAGRDTLCSLGQVAHRGMEKSLFVFTQEEVQASGLQERDMQLGFL--RALPE  503

Qy  221 -GVLETEVTYQFIDQSFQEFLAALSY.LED-----------GGVPRTAA------ 257
          || :| |  |  |  |  || |
Db  504 LGPGGDQQSYEFFHLTLQAFFTAFFLVLDDRVGTQELLRFFQEWMPPAGAATTSCYPPFL 563

Qy  258 ----GGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERXRDIERHFGCMVSERVKQEALRW 312
              || | | |  || |  |:|||| | :|  || || | |
Db  564 PFQCLQGSGPAREDLFKNKDHFQFTNLFLCGLLSKAKQK-LLRHLVPAAALRRKRKAL-W 621

Qy  313 -------VQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEEPNYPLELLYC.YETQED- 365
                     ||:  |  |  ||
Db  622 AHLFSSLRGYLKSLPRVQVESFNQVQAM---------------PTF-IWMLRCIYETQSQK 666

Qy  366 --AFVRQA.CRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKS 423
            | | | ||| | |  |  |  || ||
Db  667 VGQLAARGICA-NYLKLTYCNACSADCSALSFVLHHFPKRLA.LDLNNNL----------- 715

Qy  424 LGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSE 483
          :  :|  ||  | |:  :|  :|
Db  716 ----------NDYGVRE------LQPCFSRLT-------VLRLSVNQITDGGVKVLSE  750

Qy  484 ALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD--PQRGLQY!---- 537
          | | |  ||| | |  :  ||  | | | |  ||
Db  751 ELTKYKIVTYLGLYNNQITDVGARYVTKIL--DECKGLT-HLKLGKNITSEGGKYLALA  807

Qy  538 --------------VGM--------------LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQ 572
                          |||
Db  808 VKNSKSISEVGMWGNQVGDEGAKAFAEALRNHPSLTTLSLASNGISTEGGKSLARALQ-Q 866

Qy  573 GCGLQTLSLASVELSE--QSLQELQAVKR 599
          |||| |  |  | | ::  |||  || |  ||
Db  867 NTSLEILWLTQNELNDEVAESLAEMLKVNQ 896
```

RESULT 4
US-09-207-359B-8
; Sequence 8, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0

## Right column (RESULT 5)

```
; SEQ ID NO 8
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-207-359B-8

Query Match      9.0%; Score 289; DB 4; Length 953;
Best Local Similarity   24.2%; Pred. No. 7.1e-22;
Matches  167; Conservative  91; Mismatches 276; Indels 156; Gaps 27;

Qy    1 MLAQPQRLLFILDGADELPA---LGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLL  57
          ||||| || || | || : | |||  | |  |: ::| |   || || | | |||| |
Db  272 LLRFPHVALFTFDGLDELHSDLDLSRVPDSSC--PWEPAHPLVLLANLLSGKLLKGASKL 329

Qy   58 VTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFA 117
          | || |  |: ||::|| | |  |  :|  | :  || || | |||
Db  330 LTART--GIEVPRQFLRKKVLLRGFSPSHLRAYARRMFPERALQDRLLSQLEANPNLCS  386

Qy  118 LCFVPFVCWIVCTV--LRQQLELGRDLSRTSKTTSVLLFITSV-------L 161
          || ||| ||| :  | |  || |  |  |||
Db  387 LCSVPLFCWIIFRCFQHFRAAFEGSPQLPDCTMTLTDVFLL--VTEVHLNRMQFSSLVQRN 445

Qy  162 SSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELEQLE.RGSKVQTLFLSKKELP- 220
          ||| | | || | | |  | :| |  |  || :  ||:  |
Db  446 TRSPVETLHAGRDTLCSLGQVAHRGMEKSLFVFTQEEVQASGLQERDMQLGFL--RALPE  503

Qy  221 -GVLETEVTYQFIDQSFQEFLAALSY-LED-----------GGVPRTAA------ 257
          || :| |  |  |  |  || |
Db  504 LGPGGDQQSYEFFHLTLQAFFTAFFLVLDDRVGTQELLRFFQEWMPPAGAATTSCYPPFL 563

Qy  258 ----GGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERXRDIERHFGCMVSERVKQEALRW 312
              || | | |  || |  |:|||| | :|  || || | |
Db  564 PFQCLQGSGPAREDLFKNKDHFQFTNLFLCGLLSKAKQK-LLRHLVPAAALRRKRKAL-W 621

Qy  313 -------VQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEEPNYPLELLYC-YETQED- 365
                     ||:  |  |  ||
Db  622 AHLFSSLRGYLKSLPRVQVESFNQVQAM---------------PTF-IWMLRCIYETQSQK 666

Qy  366 --AFVRQA-CRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKS 423
            | | | ||| | |  |  |  || ||
Db  667 VGQLAARGICA-NYLKLTYCNACSADCSALSFVLHHFPKRLA-LDLNNNL----------- 715

Qy  424 LGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSE 483
          :  :|  ||  | |:  :|  :|
Db  716 ----------NDYGVRE------LQPCFSRLT-------VLRLSVNQITDGGVKVLSE  750

Qy  484 ALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD--PQRGLQY----- 537
          | | |  ||| | |  :  ||  | | | |  ||
Db  751 ELTKYKIVTYLGLYNNQITDVGARYVTKIL--DECKGLT-HLKLGKNITSEGGKYLALA  807

Qy  538 --------------VGM--------------LRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQ 572
                          |||
Db  808 VKNSKSISEVGMWGNQVGDEGAKAFAEALRNHPSLTTLSLASNGISTEGGKSLARALQ-Q 866

Qy  573 GCGLQTLSLASVELSE--QSLQELQAVKR 599
          |||| |  |  | | ::  |||  || |  ||
Db  867 NTSLEILWLTQNELNDEVAESLAEMLKVNQ 896
```

RESULT 5
US-09-340-620-8
; Sequence 8, Application US/09340620A
; Patent No. 6482933
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THE!
; FILE REFERENCE: 07334-124001
; CURRENT APPLICATION NUMBER: US/09/340,620A
; CURRENT FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/245,281
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1998-12-08

; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 8
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-340-620A-8

Query Match            9.0%;    Score 289; DB 4;  Length 953;
Best Local Similarity  24.2%;   Pred. No. 7.1e-22;
Matches 167; Conservative  91; Mismatches 276; Indels 156; Gaps 27;

```
Qy    1 MLAQPQRLLFILDGADELPA---GGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLL 57
Db  272 LLRFPHVALFTFDGLDE.HSDLDLSRVPDSSC--PWEPAHPLVLLANLLSGRLLKGASKL 329

Qy   58 V.TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENET.FA 117
Db  330 LTART--GIEVPRQFLRKKVLLRGFSPSHLRAYARRMFPERALQDRLLSQLRANPNLCS  386

Qy  118 .CFVFVCWIVCTV--LRQQ.ELGRD.SRTSKT.TSV.LLFITSV-------------L 161
Db  387 LCSVPLFCWIIFRCFQHFRAAFEGSPQLPDCTMILTDVFLL-VTEVHLRRMQPSSIVQRN 445

Qy  162 SSAPVADGPRLQGD.RNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKE.P- 220
Db  446 TRSPVETLHAGRDTLCSLGQVAHRGMEKSLFVFTQEEVQASG.QERDMQLGFL--RALPE 503

Qy  221 -GV.TETVTQFIDQSFCEFLAALSYLLED------------------GGVPR.AA--- 257
Db  504 LGPGGDQQSYEFFHLTLQAFF:AFFLVLDDRVGTQELLRFFQEMPPAGAATTSCYPPFL 563

Qy  258 -----GGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRCIERHFGCMVSERVKQEALRW 312
Db  564 PFQCLQGSGPAREDLFKNKDHFQFTNFLFLCGLLSKAKQK-LLRHLVPAAALRRKRKAL-W 621

Qy  313 -----VQGQGQGCPGVAPEVTEGAKGLEDTEEPEBEEGEEPNYPLELLYCLYETQED- 365
Db  622 AH.FSSLRGYLKSLPRVQVESFNQVQAM--------------PTF-IWMLRCIYETQSQK 666

Qy  366 --AFVRQALCFPE.ALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKS 423
Db  667 VGQLAARGICA-NYLKLTYCNACSADCSA.SFV.LHHFPKRLALD.DNNNL------- 725

Qy  424 LGKRLQASLGGGSSQGT.KQLPASLLHPLFQAMTDPLCH.SSLT.SHCKLPDAVCRDLSE 483
Db  726 --NDYGVRE------LQPCFSRLT------VLR.SVNQITDGGVKVLSE          750

Qy  484 ALRAAPALTELGLLHNRLSEAGLRMLSEGLRMLSEGLAWPQCRVQTVRVQLPD--PQRG.QYL- 537
Db  751 ELTKYKIVTYIGLYNNQITDVGARYVTKIL--DECKGL.HLKLGKNKITSEGGKYLALA 807

Qy  538 -----VGM--------------------LRQSPAL.TLDLSGCQLPAPMVTYLCAV.QHC 572
Db  808 VKNSKSISEVGMWGNQVGDEGAKAFAEALRNHPSLTTLSLASNGISTEGGKS.ARALQ-Q 866

Qy  573 GGGLCTLSLASVELSE--QSLQELQAVKR 539
Db  867 NTSLEILWLTQNELNDEVAESLAEMLKVNQ 896
```

; FILE REFERENCE: 07334/118C01
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Homo sapiens
US-09-245-281-43

Query Match            7.8%;    Score 251.5; DB 4;  Length 953;
Best Local Similarity  22.2%;   Pred. No. 8.8e-19;
Matches 164; Conservative  87; Mismatches 298; Indels 191; Gaps 27;

```
Qy    1 MLAQPQRLLFILDGADELPALPALGGPEAAP--CTDPFEAASGARVLGGLLSKALLPTALLLV 58
Db  272 LLRFPHTALFTFDG.DELHSDFDLSRVPDSCC--PWEPAHPLVLLANLLSGRLLKGAGKL  330

Qy   59 .TRAA--APGRLQGRLCSPQCAEVRGFGSDKDKKKYFYKFFRDERRAERAYRFVKENETLF 116
Db  331 .ARTGVEVPRQLRKK---VLLRGFSPSHLRAYARRMFPERTAQEH.LQQLDANPNLC   385

Qy  117 ALCFVPFVCWIV--CTVLRQQLEIG--RDLSRTSKTTTSVYLLFITSVLLSSAPVADGPRL 172
Db  386 SLCGVPLFCWIIFRCFQHFQTVFEGSSSQLPDCAVTLTDVFLL-VTEVHLNRP-QPSSLV 443

Qy  173 QGDLRNLCRLAREG----VLGRRAQ---------------FAEKE.EQLERGSKVQTLFLSKKE 218
Db  444 QRNTRSPAETLRAGWRTLHALGEVAHRGTDKSLFVFGGQEVQASKLQSGD.QLGFL--RA 501

Qy  219 LPGV-LETEVTYQFIDQSFQEFLAALSVLLEDGGV------------------------PR.AA- 257
Db  502 LPDVGPEQQGSYEFFH.LTLQAFFTAF-FLVADDKVS.RELLRFFREWTSPGEATSSSCHS 560

Qy  258 -------GGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEA 309
Db  561 SFFSFQCLGGRSRLGPDFPRNKDHFQFTNLFVCGLLAKARQKLLRQLVPKAILRR-KRKA 619

Qy  310 .RWVQGQGQGCPGVAPEVTEGAKGLEDTEPEBEEGEEGEEPNYPLELLYCLYETQEDAFVR 369
Db  620 L-WAH---------LFASLRSYLKSLPRVQSGGGFNQVHAMPTF-LWMLRCIYETQSCKVGR 669

Qy  370 QALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQ 429
Db  670 LA-----------ARGISADY.KAFCNACSADCSALSFVL------------------HHFH 703

Qy  430 ASLGGGSSQGTKQLPASLLHPLFQAMTDPLCH.SSLT.SHCKLPDAVCRDLSEALRAAP 489
Db  704 RQLALDLDNNNLNDYGVQELQPCFSRLT------VIRLSVNQIT.DTGVKVLCEELTYK 756

Qy  490 ALTELGLLHNRLSEAGLRMLSEG.AWPQCRVQTVRVQLPDPQRGLQYL-------------537
Db  757 IVTFLGLYNNQITDIGARYVA-------------QILDECRGLKH.KLGKNRITSEGG 801

Qy  538 -----VGM--------------------VGM-----KCVALAVKNSTSIVDVGMWGNQICDEGAKAFAEALKDHPSLTTLSLAFNGISPEGGKSLA 566
Db  802 KCVALAVKNSTSIVDVGMWGNQICDEGAKAFAEALKDHPSLTTLSLAFNGISPEGGKSLA 861

Qy  567 AVLQHQGCGLQTLSLASVELSEQSLQ--------------ELQAVKRA 600
Db  862 QALK-QNTTLTVIWLTKNELNDESAECFAEMLRVNQTLRHLWLIQNRITAKGTAQLARAL 920

Qy  601 KPELVI.HPALDGHPQPPKE 620
Db  921 QKNTAITEICLNGNLIKPEE 940
```

RESULT 6
US-09-245-281-43
; Sequence 43, Application US/0924528:
; Patent No. 6363196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF

RESULT 7
US-09-207-359B-43
; Sequence 43, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-207-359B-43

Query Match       7.8%;    Score 251.5;    DB 4;    Length 953;
Best Local Similarity   22.2%;   Pred. No. 8.8e-18;
Matches 164;   Conservative  87;   Mismatches  298;   Indels  191;   Gaps   27;

Qy     1 MLAQPQRLLFILDGADELPALPGPEAAP--CTDPFEAASGARVLGGLLSKALLPTALLLV  58
           |:|  :|||:|:|||    :| :|||  |    : :   |:|::|| ||:|
Db   272 LLRFPHTALFTFDGLDELHSDFDLSRVPDSCC-PWEPAHPLVLANLLSGRLLKGAGKLL  330

Qy    59 TTRAA--APGRLQGRLCSPQCAEVRGFSDKDKKKYYKFFRDERRAERAYRFVKENETLF  116
          ||||     |:|:   |:|:|    ::|||||
Db   331 TARTGVEVPRQLLRKK----VLLRGFSPSHLRAYARRMFPERTAQEHLLQQLDANPNLC  385

Qy   117 ALCFVPFVCWIV--CTVLRQQLELG--RDLSRTSKTTSVYLFITSVLSSAPVADGPRL  172
           :|   |  |:    ||| :::   | ::|:     :| :|| |  : :|:| :|:
Db   386 SLCGVPLFCWIIFRCFQHFQTVFEGSSSQLPDCAVTLTDVFLL-VTEVHLNRP-QFSSLV  443

Qy   173 QGDLRNLCRLAREG------VLGRRAQ-------------FAEKE:EQLELRGSKVQTLFLSKKE  218
           ||:|:| :|:||        |||:|                  :| | |:|:|:| :::
Db   444 QRNTRSPAETLRAGWRTLHALGEVAHRGTDKSLFVFGQEEVQASKLQEGDLQLGFL--RA  501

Qy   219 LPGV-LETEVTYQFIDQSFQEFLAALSYLLEDGGV------------VLGRRAQ  257
           ||| |:|| : | ||  | |:|||: ||:              PRTAA-----
Db   502 LPDVGPEQGQSYEFFHLTLQAFFTAF-FLVADDKVSTRELLRFFREWTSPGEATSSSCHS  560

Qy   258 --------GGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCMVSERVKQEA  309
              |:::|| | |    :| :| || : |  :|: :   :|
Db   561 SFFSFQCLGGRSRLGPDPFRNKDHFQFTNLFVCGLLAKARQKLLRQLVPKAILRR-KRKA  619

Qy   310 LRWVQGQGQGGCPGVAPEVTEGAKGLEDTEPEEEEEGEEPNYPLELLYCLYETQEDAFVR  369
           |::   | |: |  | | ||  |  |  ::| | :|| : ::|  | |: |:|:|
Db   620 L-WAH------LFASLRSYLKSLPRVQSGGFNQVHAMPTF-LWMLRCIYETQSQKVGR  669

Qy   370 QALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQ  429
            | :::  | |   |   :|  ||:   |  | | ||:|  |:: ::|:
Db   670 LA------ARGISADYLKLAFCNACSADCSALSFVL----------------HHFH  703

Qy   430 ASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP  489
           |:::|||| : ||:|:|::|| |||:| |:|:| | :|  || |:|||:|||
Db   704 RQLALDLDNNNLNDYGVQEIQPCFSRLT----------VIRLSVNQITDTGVKVLCEELTKYK  756

Qy   490 ALTELGLLHNRLSEAGLRMLSGLAWPQCRVQTVRVQLPDPQRGLQYL----  537

Qy   538 ----------------VGM-----------LRQSPALTTLDLSGCQLPAPVTYLC  566

Db   757 IVTFLGLYNNQITDIGARYVA------------QILDECRGLKHLKLGKNRITSEGG  801

Qy   567 AVLQHQGCGLQTLSLASVELSEQSLQ------------ELQAVKRA  600

Db   802 KCVALAVKNSTSIVDVGMWGNQIGDEGAKAFAEALKDHPSLTTLS-AFNGISPEGGKSLA  861

---

RESULT 8
US-09-340-620A-43
; Sequence 43, Application US/09340620A
; Patent No. 6482933
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THE
; FILE REFERENCE: 07334-124001
; CURRENT APPLICATION NUMBER: US/09/340,620A
; CURRENT FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/245,281
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 71
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 43
; LENGTH: 953
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-340-620A-43

Query Match       7.8%;    Score 251.5;    DB 4;    Length 953;
Best Local Similarity   22.2%;   Pred. No. 8.8e-18;
Matches 164;   Conservative  87;   Mismatches  298;   Indels  191;   Gaps   27;

Qy     1 MLAQPQRLLFILDGADELPALPGPEAAP--CTDPFEAASGARVLGGLLSKALLPTALLLV  58
Db   272 LLRFPHTALFTFDGLDELHSDFDLSRVPDSCC-PWEPAHPLVLANLLSGRLLKGAGKLL  330

Qy    59 TTRAA--APGRLQGRLCSPQCAEVRGFSDKDKKKYYKFFRDERRAERAYRFVKENETLF  116
Db   331 TARTGVEVPRQLLRKK----VLLRGFSPSHLRAYARRMFPERTAQEHLLQQLDANPNLC  385

Qy   117 ALCFVPFVCWIV--CTVLRQQLELG--RDLSRTSKTTSVYLFITSVLSSAPVADGPRL  172
Db   386 SLCGVPLFCWIIFRCFQHFQTVFEGSSSQLPDCAVTLTDVFLL-VTEVHLNRP-QPSSLV  443

Qy   173 QGDLRNLCRLAREG------VLGRRAQ-------------FAEKELEQLELRGSKVQTLFLSKKE  218
Db   444 QRNTRSPAETLRAGWRTLHALGEVAHRGTDKSLFVFGQEEVQASKLQEGDLQLGFL--RA  501

Qy   219 LPGV-LETEVTYQFIDQSFQEFLAALSYLLEDGGV------------PRTAA-----  257
Db   502 LPDVGPEQGQSYEFFHLTLQAFFTAF-FLVADDKVSTRELLRFFREWTSPGEATSSSCHS  560

Qy   258 --------GGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCMVSERVKQEA  309
Db   561 SFFSFQCLGGRSRLGPDPFRNKDHFQFTNLFVCGLLAKARQKLLRQLVPKAILRR-KRKA  619

Qy   310 LRWVQGQGQGGCPGVAPEVTEGAKGLEDTEPEEEEEGEEPNYPLELLYCLYETQEDAFVR  369
Db   620 L-WAH------LFASLRSYLKSLPRVQSGGFNQVHAMPTF-LWMLRCIYETQSQKVGR  669

Qy   370 QALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQ  429
Db   670 LA------ARGISADYLKLAFCNACSADCSALSFVL----------------HHFH  703

Qy   430 ASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAP  489
Db   704 RQLALDLDNNNLNDYGVQEIQPCFSRLT----------VIRLSVNQITDTGVKVLCEELTKYK  756

Qy   490  ALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYL---------      537
          |.| || :|||:||  ::| |.::        .||:||.|:||.|:|
Db   757  IVTFLGLYNNQITDIGARYVA-----------QILDECRGLKHLKLGKNRITSEGG       801

Qy   538  -------------VGM----------------------LRQSPAITTLDLSGCQLPAPMVTYLC      566
                        |||                      |||||||||||||||||||||||||||
Db   802  KCVALAVKNSTSIVDVGMWGNQIGDEGAKAFAEALKDHPSLTTLSLAFNGISPEGGKSLA      861

Qy   567  AVLQHQGGLQTLSLASVELSEQSLQ-------------------ELQAVKRA      620
          :.|   |: :|.|:|| | ||| |                    |:||||.
Db   862  QALK-QNTTLVTWLTKQELNDESAECFAEMLRVNQTLRHLMLIQNRITAKGTAQLARAL      920

Qy   601  KPDLVTHPALDSHPQQPPKE      620
          .| | :   | ||: ||
Db   921  QXNTAITE-CLNGNLIKPEE      940

RESULT 9
US-09-207-359B-47
; Sequence 47, Application US/0920'359B
; Patent No. 6469140
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;   TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;   FILE REFERENCE: 07334-112001
;   CURRENT APPLICATION NUMBER: US/09/207,359B
;   CURRENT FILING DATE: 1998-12-08
;   PRIOR APPLICATION NUMBER: US 39/099,041
;   PRIOR FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: US 09/019,942
;   PRIOR FILING DATE: 1998-02-06
;   NUMBER OF SEQ ID NOS: 47
;   SEQ ID NO 47
;   LENGTH: 966
;   TYPE: PRT
;   ORGANISM: Artificial Sequence
;   FEATURE:
;   OTHER INFORMATION: majority sequence
US-09-207-359B-47

Query Match           7.0%;  Score 224.5;  DB 4;  Length 966;
Best Local Similarity  23.1%;  Pred. No. 7.9e-15;
Matches 161;  Conservative 84;  Mismatches 296;  Indels 155;  Gaps 30;

Qy     1  MLAQPQRLLFILDGADELPA---LGGPEAAPCTDPEAASGARVLGGLLSKALLPTALL.      57
          : |  | .||. .| | ||    .||| |: :|  |.||| :||:||.|..| |.||
Db   272  LJRFPHVALFTFDGLDELHSDLDLSRVPDSSC--PWEPAHPLVLLANLLSGKLLKGAGKL      329

Qy    58  VTTRAA--APGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETL      115
          :.| ||    || |.|   .:||    .|||        |:|||.|||.|||.|:|.|.|
Db   330  LTARTGVEVPRQLLRKK-----VLRGFSPSHLRAYARRMFPERAAQDHLLSQLDANFVL      384

Qy   116  FALCFVPFVCWIV--CTVLRQQLELG--RDLSRTSKTTTSVYLL---FITSVLSSAFVA      167
          || ||:|   | |  ||.|.|. || |  :|. ||  | |::       .||:|||. .
Db   385  CSLCGPVLFCWIIFRCFQQHFQAAFEGSSSQLPDCAVTLTDVFLLVTEVHLNRMQPSSLVQ      444

Qy   168  DGPRLQGD-----------LRNLCRLAREGVLGRRAQFAEKELEQ:ELRGSKVQTLFLSKKEL      219
          .||  ||              :| |.|:|| | ||  | ||. | ||::|.|.| |||.||
Db   445  RNTRSPAETLHAGRETLHALGEVAHRGTDKS:FVFGQEEVQASGLQEGLQLGFL--RAL      502

Qy   220  P--GVLETEVTYQFIDQSFQEFLAALSYLLED---------------------GGVPRTAA---      257
          |   |:.|| .|.|.|| |:.:   | |  |                       ||.| ||
Db   503  PDVGPGGDGSYEFFHLTLQAFF:AFFLVADDXVGTQELLRFFQEWTSPGGAASSCHSS      562

Qy   258  ------GGVGTLLRGDAQFASHLVLTRFLFGLLSAERMRDIERHFGCMVSERVKQEAL      310
               .: |.|::| :|||:.|:||.|.||||:| :|:||:| ||||.:|.||||||.
Db   563  FLSFCCLGGSGRAGEDLFKNYDHFQFTNLFVCGLL-AKAKQKLLRQLVPAAALRRKRXAI      621

Qy   311  RWVQ--GQQGGCPGVAPEVTEGAKGLEDTEEPEEEEEGEEPNYPLELLYCLYETQEDAFV      368
          : .:    |:|.|| .||:.|   ||.||||:|||||||||||||:|||:||:|||:|:|

Best Local Similarity   27.4%;   Pred. No. 1.9e-14;
Matches   77;  Conservative   45;  Mismatches   113;  Indels   45;  Gaps   7;

```
Qy   354  ELLYCLYETQ---EDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAG------ 403
                 |||   : ::   :|  :|   :: :||| |  :|  :  :|  ||
Db    17  ELLPLLQQYEVVRLDDCGLTEEHCKDIGSAL-------RANPSLTELCLRTNELGDAGVHL 70

Qy   404  --QALRLISCRIVAAQEKKKKS:GKRLQASLGGSSQGTTKQLPA----------- 448
             |:||:::  ::  |  :    |   |:  |:
Db    71  VLQGLQSPTCKI------QKLSLQNCSLTEAGCGVLPSTLRSLPTLRELHLSDNPLGDAG 124

Qy   449  LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRM 508
             | | | : |:|: ::|  ||| :| :|  |::  | || ::|  |  :
Db   125  LRLLCEGLLDPQCHLEKLQLEYCRLTAASCEPLASVLRATRALKELTVSNNDIGEAGARV 184

Qy   509  LSEGLAWPQCRVQTVRVQLPDPQRGL-----QYLVGMLRQSPALTTLDLSGCQLPAPMVT 563
             |::|| |  || |  |  |        |  |:|  |: ::|  :| :|
Db   185  LGQGLADSACQLETLRLE---NCGLTPANCKDLCGIVASQASLRELDLGSNGLGDAGIA 240

Qy   564  YLCAVLQHGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
             | :| :||| | : |: ::  ::   | | :|||:
Db   241  ELCPGLLSPASRLKTLWLWECDITASGCRDLCRVLQAKETL 281
```

RESULT 11
US-08-795-395-2
; Sequence 2, Application US/08795395
; Patent No. 5965399
; GENERAL INFORMATION:
;  APPLICANT: CHATTERJEE, DEB K.
;  APPLICANT: SHANDILYA, HARINI
;  TITLE OF INVENTION: Cloning and Expression of Rat Liver and
;  TITLE OF INVENTION: Porcine Liver Ribonuclease Inhibitor
;  NUMBER OF SEQUENCES: 8
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;   STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;   CITY: WASHINGTON
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3934
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/795,395
;   FILING DATE: 04-FEB-1997
;   CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/024,057
;   FILING DATE: 16-AUG-1996
;  ATTORNEY/AGENT INFORMATION:
;   NAME: ESMOND, ROBERT W.
;   REGISTRATION NUMBER: 32,893
;   REFERENCE/DOCKET NUMBER: 0942.3440002
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-2600
;   TELEFAX: 202-371-2540
;  INFORMATION FOR SEQ ID NO: 2:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 456 amino acids
;    TYPE: amino acid
;    TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-795-395-2

Query Match    6.7%;  Score 216;  DB 2;  Length 456;
Best Local Similarity   27.4%;  Pred. No. 1.9e-14;
Matches   77;  Conservative   45;  Mismatches   113;  Indels   45;  Gaps   7;

Qy   354  ELLYCLYETQ---EDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAG------ 403

---

```
Db    17  ELLPLLQQYEVVRLDDCGLTEEHCKDIGSAL-------RANPSLTELCLRTNELGDAGVHL 70

Qy   404  --QALRLISCRIVAAQEKKKKSLGKRLQASLGGSSQGTTKQLPA----------- 448
Db    71  VLQGLQSPTCKI------QKLSLQNCSLTEAGCGVLPSTLRSLPTLRELHLSDNPLGDAG 124

Qy   449  LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRM 508
Db   125  LRLLCEGLLDPQCHLEKLQLEYCRLTAASCEPLASVLRATRALKELTVSNNDIGEAGARV 184

Qy   509  LSEGLAWPQCRVQTVRVQLPDPQRGL-----QYLVGMLRQSPALTTLDLSGCQLPAPMVT 563
Db   185  LGQGLADSACQLETLRLE---NCGLTPANCKDLCGIVASQASLRELDLGSNGLGDAGIA 240

Qy   564  YLCAVLQHGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
Db   241  ELCPGLLSPASRLKTLWLWECDITASGCRDLCRVLQAKETL 281
```

RESULT 12
US-08-910-731-8
; Sequence 8, Application US/08910731
; Patent No. 5932440
; GENERAL INFORMATION:
;  APPLICANT: CHATTERJEE, DEB K.
;  APPLICANT: SHANDILYA, HARINI
;  TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
;  NUMBER OF SEQUENCES: 16
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;   STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;   CITY: WASHINGTON
;   STATE: D.C.
;   COUNTRY: USA
;   ZIP: 20005-3934
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: PatentIn Release #1.0, Version #1.30
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/08/910,731
;   FILING DATE: (Herewith)
;   CLASSIFICATION: 435
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/795,395
;   FILING DATE: 04-FEB-1997
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 08/794,546
;   FILING DATE: 03-FEB-1997
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER: 60/024,057
;   FILING DATE: 16-AUG-1996
;  ATTORNEY/AGENT INFORMATION:
;   NAME: ESMOND, ROBERT W.
;   REGISTRATION NUMBER: 32,893
;   REFERENCE/DOCKET NUMBER: 0942.3440003
;  TELECOMMUNICATION INFORMATION:
;   TELEPHONE: 202-371-2600
;   TELEFAX: 202-371-2540
;  INFORMATION FOR SEQ ID NO: 8:
;   SEQUENCE CHARACTERISTICS:
;    LENGTH: 456 amino acids
;    TYPE: amino acid
;    STRANDEDNESS: both
;    TOPOLOGY: both
;   MOLECULE TYPE: peptide
US-08-910-731-8

Query Match    6.7%;  Score 214;  DB 2;  Length 456;
Best Local Similarity   27.4%;  Pred. No. 3.1e-14;
Matches   77;  Conservative   45;  Mismatches   113;  Indels   46;  Gaps   7;

Best Local Similarity 29.2%; Pred. No. 5.2e-14;
Matches 59; Conservative 35; Mismatches 104; Indels 4; Gaps 2;

```
Qy  404 CALRLISCRLVAAQEKKKSLSGKRLQASLGGGSSQGTTKQLPASLLHPLFCAMTDPLCHL 463
            ::|: :|||  :|::|:::| ::||:|:|:|::|
Db   31 QVVRLDDCGL?---EARCKDISSALRVNPALAELNLRSNELGDVGVHCVLQGLQTPSCKI 87

Qy  464 SSLT:SHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSPAGLRMLSEGLAWPQCRVQTV 523
            ::||:|||:|:|||||||||:||||||||:|||:||||
Db   88 QKLSLQNCCLTGAGCGVLSSTLRTLPTLPLQELHLSDNLLGDAGLQLLCEGLLDPQCRLEKL 147

Qy  524 RVQLPD-PQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLA 582
            :::    :|:|::||:|::|||:|||:||
Db  148 QLEYCSLSAASCEPLASVLRAKPDFKELTVSNNDINEAGVRVLCQGLKDSPCCLEALKLE 207

Qy  583 SVELSEQSLQELQAVKRAKPDL 604
            :::
Db  208 SCGVTSDNCRDLCGIVASKASL 229
```

RESULT 14
US-08-910-731-4
; Sequence 4, Application US/08910731
; Patent No. 5932440
; GENERAL INFORMATION:
;   APPLICANT: CHATTERJEE, DEB K.
;   APPLICANT: SHANDILYA, HARINI
;   TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;     STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;     CITY: WASHINGTON
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/910,731
;     FILING DATE: (Herewith)
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/795,395
;     FILING DATE: 04-FEB-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/794,546
;     FILING DATE: 03-FEB-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/024,057
;     FILING DATE: 16-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: ESMOND, ROBERT W.
;     REGISTRATION NUMBER: 32,893
;     REFERENCE/DOCKET NUMBER: 0942.344C003
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-371-2600
;     TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 456 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-910-731-4

Query Match             6.5%;  Score 209.5;  DB 2;  Length 456;
Best Local Similarity 24.3%;  Pred. No. 9.6e-14;
Matches 93; Conservative 45; Mismatches 129; Indels 115; Gaps 15;

---

```
Qy  354 ELLYCLYETQ---EDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAG------ 403
            ::|              :|::    ::| :
Db   17 ELLPLLQQYEVVRLDDCGLTEEHCKDIGSAL-----RANPSLTELCLRTNELGDAGVHL 70

Qy  404 -CALRLISCRLVAAQEKKKSLSGKRLQASLGGGSSQGTTKQLPA---------SL 448
            :|::|  :||| :|::|:::|  ::||:|:|:
Db   71 VLQGLQSPTCKI-----QKLSLQNCSLTEAGCGVLPSTLRSLPTLRELHLSDNPLGDAG 124

Qy  449 LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRM 508
            ::|::  ::|||:||:|:|||||||||||||:|:|||||||:|||||||:
Db  125 LRLLCEGLLDPQCHLEKQLEYCRLTAASCEPLASVLRATRA.KELTVSNNDIGEAGARV 184

Qy  509 LSEGLAWPQCRVCTVRVQLPDPQRG----QYLVGMLRQSPALTTLDLSGCQLPAPMVT 563
            ::|:|   :|::|:||:|     ::|:|:|||||:|||:|
Db  185 LGQGLADSACQLETLRLE----NCGLTPANCKDLCGIVASQASLRELALGSNKLGDVGMA 240

Qy  564 YLCAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL 604
            :|:::|:||:
Db  241 ELCPGLLHPSSRLRTLWIWECGITAKGCGDLCRVLRAKFSL 281
```

RESULT 13
US-08-910-731-6
; Sequence 6, Application US/08910731
; Patent No. 5932440
; GENERAL INFORMATION:
;   APPLICANT: CHATTERJEE, DEB K
;   APPLICANT: SHANDILYA, HARINI
;   TITLE OF INVENTION: Mammalian Ribonuclease Inhibitors and Use Thereof
;   NUMBER OF SEQUENCES: 16
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX P.L.L.C.
;     STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;     CITY: WASHINGTON
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/910,731
;     FILING DATE: (Herewith)
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/795,395
;     FILING DATE: 04-FEB-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 08/794,546
;     FILING DATE: 03-FEB-1997
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/024,057
;     FILING DATE: 16-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: ESMOND, ROBERT W.
;     REGISTRATION NUMBER: 32,893
;     REFERENCE/DOCKET NUMBER: 0942.344C003
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-371-2600
;     TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 6:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 461 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: single
;     TOPOLOGY: not relevant
;   MOLECULE TYPE: peptide
US-08-910-731-6

Query Match             6.6%;  Score 212;  DB 2;  Length 461;

```
Query Match              6.5%;  Score 209.5;  DB 2;  Length 456;
Best Local Similarity   24.3%;  Pred. No. 9.6e-14;
Matches  93;  Conservative 45;  Mismatches 129;  Indels 115;  Gaps 15;

Qy 259 GVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGQ 318
           ||:|:|:|    |:|                               ::|
Db  67 GVGLVLQGLQNPTCKI----------QKLSLQ-------------NCSLTE---------A  95

Qy 319 GCPGVAPEVTEGAKGLED---TEEPEEEGEEPNYPLELLYCLYETQEDAFVRQALC--    373
       || |:    |::    |:|                        ::|
Db  96 GC-GVLPDVLRSLSTLRELHLNDNPLGDE------GLKJ---------------LCEG    131

Qy 374 -RFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKSLGKRLQASL 432
        |::|    :    :|      :|                 :|
Db 132 LRDPQ--------CRLEKLQLEYC-------NLTATSCEPLASVLRVKPDF-KELVLSN  174

Qy 433 GGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALT 492
        :       :    |:   : :|         : :|      : :|
Db 175 NDFHEAG-------IHTLCQGLKDSACQLESLKLENCGITSANCKDLCDVVASKASLQ   225

Qy 493 ELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD-PQRGLQYLVGMLRQSPALTTLD 551
       |: | ::|  | |  |  |     :|:    :|      :| :|       :|
Db 226 ELDLGSNKLGNTGIAALCSGLLLPSCRLRTLWLWDCDVTAEGCKDLCRVLRAKQSLKELS 285

Qy 552 LSGCQLPAPMVTYLCAVLQHQGCGLQTL-------------SLASVELSE          589
       |:|  |                                   :::
Db 286 LAGNELKDEGAQLLCESLLEPGCQLESLWVKTCSLTAASCPHFCSVLTKNSSLFELQMSS 345

Qy 589 QSLQELQAVKRAK----PDLVI                                      606
       :: |                :|
Db 346 NPLGDSGVVELCKALGYPDTVL                                      367

Search completed: October 2, 2003, 17:51:53
Job time : 33 secs
```

RESULT 15
US-08-795-395-4
; Sequence 4, Application US/08795395
; Patent No. 5965399
; GENERAL INFORMATION:
;   APPLICANT: CHATTERJEE, DEB K.
;   APPLICANT: SHANDILYA, HARINI
;   TITLE OF INVENTION: Cloning and Expression of Rat Liver and
;   TITLE OF INVENTION: Porcine Liver Ribonuclease Inhibitor
;   NUMBER OF SEQUENCES: 8
;   CORRESPONDENCE ADDRESS:
;     ADDRESSEE: STERNE, KESSLER, GOLDSTEIN & FOX, P.L.L.C.
;     STREET: 1100 NEW YORK AVE., N.W., SUITE 600
;     CITY: WASHINGTON
;     STATE: D.C.
;     COUNTRY: USA
;     ZIP: 20005-3934
;   COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/08/795,395
;     FILING DATE: 04-FEB-1997
;     CLASSIFICATION: 435
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: 60/024,057
;     FILING DATE: 16-AUG-1996
;   ATTORNEY/AGENT INFORMATION:
;     NAME: ESMOND, ROBERT W.
;     REGISTRATION NUMBER: 32,893
;     REFERENCE/DOCKET NUMBER: 0912.3440002
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 202-371-2600
;     TELEFAX: 202-371-2540
; INFORMATION FOR SEQ ID NO: 4:
;   SEQUENCE CHARACTERISTICS:
;     LENGTH: 456 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;   MOLECULE TYPE: protein
US-08-795-395-4

```
                  Copyright    GenCore version 5.1.6
                  Copyright (c) 1993 - 2003  Compugen Ltd.

CM protein - protein search, using sw model

Run on:    October 2, 2003, 17:35:58 ; Search time 24 Seconds
                                         (without alignments)
                                         1224.654 Million cell updates/sec

Title:     US-10-029-347-2
Perfect score: 3218
Sequence:   1 MLAQPQRLLFILDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table: BLOSUM62
               Gapop 10.0 , Gapext 0.5

Searched:   127863 seqs, 47026705 residues

Total number of hits satisfying chosen parameters:    127863

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Database :   SwissProt_41:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

                            SUMMARIES
```

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 3207.5 | 99.7 | 892 | 1 | PYA5_HUMAN | P59044 homo sapien |
| 2 | 2183 | 67.6 | 854 | 1 | PYA5_RAT | Q63035 rattus norv |
| 3 | 2227.5 | 66.1 | 843 | 1 | PYA5_MOUSE | Q91ws2 mus musculu |
| 4 | 835 | 25.9 | 1053 | 1 | CIS1_MOUSE | Q8r1b8 mus musculu |
| 5 | 825 | 25.6 | 1062 | 1 | PYA7_HUMAN | P59046 homo sapien |
| 6 | 800.5 | 24.9 | 1034 | 1 | CIS1_HUMAN | Q96p20 homo sapien |
| 7 | 762 | 23.7 | 1473 | 1 | NAL1_HUMAN | Q9c000 homo sapien |
| 8 | 650 | 20.2 | 980 | 1 | PYA3_HUMAN | Q8wx94 homo sapien |
| 9 | 616.5 | 19.2 | 1062 | 1 | NAL2_HUMAN | Q9nx02 homo sapien |
| 10 | 598.5 | 18.6 | 994 | 1 | NAL4_HUMAN | Q96mn2 homo sapien |
| 11 | 507.5 | 15.8 | 1200 | 1 | MATE_HUMAN | P59947 homo sapien |
| 12 | 468.5 | 14.6 | 1111 | 1 | MATE_MOUSE | Q9z1m5 mus musculu |
| 13 | 461 | 14.3 | 1033 | 1 | PYA6_HUMAN | P59045 homo sapien |
| 14 | 305 | 9.5 | 1020 | 1 | CARF_MOUSE | Q8k3z0 mus musculu |
| 15 | 304.5 | 9.5 | 1130 | 1 | C2TA_HUMAN | P33076 homo sapien |
| 16 | 300.5 | 9.3 | 1040 | 1 | CARF_HUMAN | Q9hc29 homo sapien |
| 17 | 289 | 9.0 | 953 | 1 | C2TA_HUMAN | Q9y239 homo sapien |
| 18 | 267.5 | 8.3 | 1155 | 1 | C2TA_MOUSE | P79621 mus musculu |
| 19 | 254.5 | 7.9 | 953 | 1 | CAR4_MOUSE | Q8bhb0 mus musculu |
| 20 | 216 | 6.7 | 456 | 1 | RINI_PIG | P10775 sus scrofa |
| 21 | 212 | 6.6 | 460 | 1 | RINI_HUMAN | P13489 homo sapien |
| 22 | 209.5 | 6.5 | 456 | 1 | RINI_RAT | P29315 rattus norv |
| 23 | 123 | 3.8 | 1024 | 1 | CARC_HUMAN | Q9npp4 homo sapien |
| 24 | 116 | 3.6 | 1403 | 1 | BIRE_MOUSE | Q9r016 mus musculu |
| 25 | 113.5 | 3.5 | 681 | 1 | SSAV_SALTY | P74856 salmonella |
| 26 | 112.5 | 3.5 | 1238 | 1 | SBCC_RHOCA | O68032 rhodobacter |
| 27 | 108.5 | 3.4 | 406 | 1 | HUTI_VIBVU | Q8da18 vibric vuln |
| 28 | 107.5 | 3.3 | 780 | 1 | PRTP_HSV1F | P06490 herpes simp |
| 29 | 107.5 | 3.3 | 3321 | 1 | PCN2_HUMAN | O95613 homo sapien |
| 30 | 106 | 3.3 | 1173 | 1 | UB4B_MOUSE | Q9es00 mus musculu |
| 31 | 104.5 | 3.2 | 1795 | 1 | ES21_HUMAN | Q14674 homo sapien |
| 32 | 104 | 3.2 | 857 | 1 | CA2P_THES7 | P51060 thermus sp. |
| 33 | 104 | 3.2 | 1403 | 1 | BIRA_MOUSE | Q9qwk5 mus musculu |
| 34 | 103 | 3.2 | 740 | 1 | DAXX_HUMAN | Q9uer7 homo sapien |
| 35 | 103 | 3.2 | 1189 | 1 | POL_BAEVM | P10272 baboon endo |
| 36 | 102.5 | 3.2 | 383 | 1 | DTC_HUMAN | P01880 homo sapien |
| 37 | 102.5 | 3.2 | 915 | 1 | SYA_METKA | Q8twy1 methanopyru |
| 38 | 102.5 | 3.2 | 1410 | 1 | RRBI_HUMAN | Q9p2e9 homo sapien |
| 39 | 102 | 3.2 | 509 | 1 | LEGA_GOSHI | P09802 gossypium h |
| 40 | 102 | 3.2 | 1148 | 1 | CNN_DROME | P54623 drosophila |
| 41 | 101.5 | 3.2 | 736 | 1 | DAXX_CERAE | O18C05 cercopithec |
| 42 | 101 | 3.1 | 3674 | 1 | SPCR_HUMAN | Q9nrc6 homo sapien |
| 43 | 98.5 | 3.1 | 1138 | 1 | RDL3_ARATH | Q8w3k0 arabidopsis |
| 44 | 98 | 3.0 | 1302 | 1 | UB4B_HUMAN | O95155 homo sapien |
| 45 | 97.5 | 3.0 | 620 | 1 | GG95_HUMAN | Q08379 homo sapien |

ALIGNMENTS

```
RESULT 1
PYA5_HUMAN
ID   PYA5_HUMAN     STANDARD;     PRT;     892 AA.
AC   P59044;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   PYRIN-containing APAF1-like protein 5.
GN   PYPAF5.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22162427; PubMed=12019269;
RA   Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA   Lora J.M., Geddes B.J., Briskin M., DiStefano P.S., Bertin J.;
RT   "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT   activation of NF-kappa B and caspase-1-dependent cytokine
RT   processing.";
RL   J. Biol. Chem. 277:29874-29880(2002).
RN   [2]
RP   FUNCTION.
RX   MEDLINE=22275822; PubMed=12387869;
RA   Grenier J.M., Wang L., Manji G.A., Huang W.-J., Al-Garawi A.,
RA   Kelly R., Carlson A., Merriam S., Lora J.M., Briskin M.,
RA   DiStefano P.S., Bertin J.;
RT   "Functional screening of five PYPAF family members identifies PYPAF5
RT   as a novel regulator of NF-kappaB and caspase-1.";
RL   FEBS Lett. 530:73-78(2002).
CC   -!- FUNCTION: May mediate activation of CASP1 via ASC and promote
CC       activation of NF-kappa-B.
CC   -!- SUBUNIT: Binds to ASC with its DAPIN domain.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Highly expressed in granulocytes. Detected at
CC       much lower levels in T-cells.
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 5 leucine-rich (LRR) repeats.
CC   ----------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ----------------------------------------------------------------------
DR   EMBL; AF479748; AAL87105.1; -.
DR   InterPro; IPR007091; LRR_RNinh;.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   SMART; SM00368; LRR_RI; 3.
```

```
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat.
FT   DOMAIN        1    103       DAPIN.
FT   DOMAIN      196    513       NACHT.
FT   DOMAIN      604    614       POLY-GLU.
FT   REPEAT      462    487       LRR 1.
FT   REPEAT      727    747       LRR 2.
FT   REPEAT      755    778       LRR 3.
FT   REPEAT      811    834       LRR 4.
FT   REPEAT      845    868       LRR 5.
FT   NP_BIND     202    209       ATP (POTENTIAL).
SQ   SEQUENCE   892 AA;  98733 MW;  4AA8D1FC766DDE9D CRC64;

Query Match             99.7%;  Score 3207.5;  DB 1;  Length 892;
Best Local Similarity   99.8%;  Pred. No. 1.3e-226;
Matches  625;  Conservative   0;  Mismatches   0;  Indels   1;  Gaps   1;

Qy       1 MLAQPQRLFLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 60
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     267 MLAQPQRLFLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 326

Qy      61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETIFALCF 120
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     327 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETIFALCF 386

Qy     121 VPFVCWIVCTVLRQQLEGRDLSRFSKTTTSVYLFITSVLSSAPVADGPRLQGDLRNLC 180
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     387 VPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLFITSVLSSAPVADGPRLQGDLRNLC 446

Qy     181 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 240
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     447 RLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEFL 506

Qy     241 AALSYLLEDGGVPRTAAGGVTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCM 300
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     507 AALSYLLEDGGVPRTAAGGVTLLRGDAQPHSHLVLTTRFLFGLLSAERXRDIERHFGCM 566

Qy     301 VSERVKQEALRWVQQGQGQCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELLYCLY 360
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     567 VSERVKQEALRWVQQGQGQCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELLYCLY 626

Qy     361 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     627 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKK 686

Qy     421 KKSLGKRLQASLGGG-SSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCR 479
           ||||||||||||||| ||||||||||||||||||||||||||||||||||||||||||||
Db     687 KKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCR 746

Qy     480 DLSEALRAAPALTELGLJHNRLSEAGLRML-SEGLAWPQCRVQTVRVQLPDPQRGLQYLVG 539
           ||||||||||||||||||||||||||||||| ||||||||||||||||||||||||||||
Db     747 DLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVG 806

Qy     540 MLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHGGCGLQTLSLASVELSEQSLQELQAVKR 599
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     807 MLRQSPALTLDLSGCQLPAPMVTYLCAVLCAVLQHGGCGLQTLSLASVELSEQSLQELQAVR 866

Qy     600 AKPDLVITHPALDGHPQPPKELISTF 625
           ||||||||||||||||||||||||||
Db     867 AKPDLVITHPALDGHPQPPKELISTF 892


RESULT 2
PYA5_RAT
ID   PYA5_RAT      STANDARD;      PRT;   854 AA.
AC   Q63C35;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   PYRIN-containing APAF1-like protein 5-like (Angiotensin II/vasopressin
DE   receptor).
GN   PYPAF5 OR AVR.
```

```
OS   Rattus norvegicus (Rat).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
OX   NCBI_TaxID=10116;
RN   [1]
RP   CONCEPTUAL TRANSLATION OF 1-343.
RA   Hinz U.;
RL   Unpublished observations (FEB-2003).
RN   [2]
RP   SEQUENCE OF 344-854 FROM N.A., MUTAGENESIS OF GLU-393, AND FUNCTION.
RC   TISSUE=Kidney;
RX   MEDLINE=96071640; PubMed=7489366;
RA   Ruiz-Opazo N., Akimoto K., Herrera V.L.M.;
RT   "Identification of a novel dual angiotensin II/vasopressin receptor on
RT   the basis of molecular recognition theory.";
RL   Nat. Med. 1:1074-1081(1995).
RN   [3]
RP   VARIANTS SALT-SENSITIVE HYPERTENSION SER-492 AND ARG-536.
RX   MEDLINE=21980185; PubMed=11984003;
RA   Ruiz-Opazo N., Lopez L.V., Herrera V.L.M.;
RT   "The dual AngII/AVP receptor gene N119S/C163R variant exhibits
RT   sodium-induced dysfunction and cosegregates with salt-sensitive
RT   hypertension in the Dahl salt-sensitive hypertensive rat model.";
RL   Mol. Med. 8:24-32(2002).
RN   [4]
RP   IDENTIFICATION OF MAMMALIAN ORTHOLOGS OF PYPAF5.
RA   Albrecht M., Domingues F.S., Schreiber S., Lengauer T.;
RT   "Identification of mammalian orthologs associates PYPAF5 with distinct
RT   functional roles.";
RL   FEBS Lett. 538:173-177(2003).
CC   -!- FUNCTION: May mediate activation of CASP: via ASC and promote
CC       activation of NF-kappa-B (By similarity). Angiotensin II and
CC       vasopressin binding protein. May stimulate cAMP accumulation.
CC   -!- SUBUNIT: Binds to ASC with its DAPIN domain (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- TISSUE SPECIFICITY: Highly expressed in kidney. Detected at low
CC       levels in all tissues tested.
CC   -!- DISEASE: Defects in PYPAF5 may be a cause of salt-sensitive
CC       hypertension.
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 4 leucine-rich (LRR) repeats.
CC   -!- CAUTION: The N-terminus was extended using ESTs and genomic
CC       sequences, in analogy to ortholog sequences.
CC   ---------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   ---------------------------------------------------------------------------
DR   EMBL; M85183; AAA03623.1; ALT_INIT.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat.
FT   DOMAIN        1    103       DAPIN.
FT   DOMAIN      168    484       NACHT.
FT   DOMAIN      564    604       ASP/GLU-RICH.
FT   DOMAIN      655    662       POLY-LYS.
FT   REPEAT      433    458       LRR 1.
FT   REPEAT      610    633       LRR 2.
FT   REPEAT      722    745       LRR 3.
FT   REPEAT      812    836       LRR 4.
FT   NP_BIND     174    181       ATP (POTENTIAL).
FT   VARIANT     492    492       N -> S (IN SALT-SENSITIVE HYPERTENSION).
FT   VARIANT     536    536       C -> R (IN SALT-SENSITIVE HYPERTENSION).
FT   MUTAGEN     766    766       E->K: ABOLISHES ANGIOTENSIN II BINDING.
```

SQ  SEQUENCE  854 AA;  95292 MW;  D7B8D922D77B3734 CRC64;

Query Match                  67.8%;  Score 2183;  Pred. No. 8.6e-152;  DB 1;  Length 854;
Best Local Similarity        71.0%;
Matches 443;  Conservative 55; Mismatches 112; Indels 14; Gaps 5;

Qy      1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
          ::|| |||||||||||||||| ||||||||||||:| :|||||||| ||:|| ||||||||
Db    239 ILAQPHRLLFILDGADELPTIAAPEATPCRDPFEATSGLRVLSGLLSQELLPSARLLVTS 298

Qy     61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
          ||| ||||||||||||||||||||||||||||||||||:||||||||||||||||||||||
Db    299 RNATLGRLQGRLCSPQCAEVRGFSDKDKKKYFFKFFRDERKAERAYRFVKENETLYALCF 358

Qy    121 VPFVCWIVCTVLRQQLELGRGLSRTSKTTTSVYLFITSVLSSAPVADGPRLQGDLRNLC 180
          |||||||||| ||||| | ||| ||:| |||| ||||:||||| |: ||| ||| ||| :|
Db    359 VPFVCWIVCTVLLQQMELGRDLSRTSKTTTSVYLFITSMLKSAG-TNGPRVQGELRMLC 417

Qy    181 RLAREGVLGRRAQFAEKELEOLELRGSKVQTLFLSKKELPGV-ETEVTYQFIDQSFQEFL 240
          |||||| | ||||||| |:|||| |||| || || |||||||:|||||||||||||||||
Db    418 RLAREGILKGHQAQFSEKDLERLKLQGSQVQTMFLSKKELPGVLETVVTYQFIDQSFQEFL 477

Qy    241 AALSYLLEDGGVPRTAACGVGTLLRGDAQPHSELVLTTRFLFGLLSAERMRDIERHFGCM 300
          |||| ||:|||| | | ||::| ||| |||: |||||||||||||:|||| ||:|||||
Db    478 AALSYLLDAEGAPGNSAGSVCMLLNSDAGLRGHLALTTRFLFGLLSTERIRDIGNHFGCV 537

Qy    301 VSERVKCEALRWVCGQGGCPGVA--PEVTEGAKGEDTEPEEEGEERNYPLELLY 357
          |:|| |||:| |||| |||||   :||| |||||:||||||||| ||:|||||||
Db    538 VPGRVKQDTLRWVQGQSQ--PKVATVGAEKKDELKDEEAEEEEEEENFGLELLY 595

Qy    358 CLYETQEDDAFVRQALCRFPELAIQRVRFCRMDVAVLSYCVRCCPAGQALRLSCRLVAAQ 417
          |||||||||:|||| || | |:| |: || |||:||| | : :|:|| |||: ||||:|
Db    596 CLYETQEDDFVRQALSSLPEMVLERVRLTRMDLEVLSYCVQCCPDGQALRJVSCGLVAAK 655

Qy    418 E-KKKKSLGKRLQASLGGGSSOGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPD 475
            : || | |||:|| ::: |:| | :||| |||::||:||| ||| |:||:|| ||| :|
Db    656 EKKKKKSFMNRLK------GSQSTGKQPPASLLRPLCEAMITQQCGLSILTLSHCKLPC 709

Qy    476 AVCRDLSEALRAAPALTELGLLNRLSEAGLRMLSEGLAWPQCRVOTVRVQLPDPQRGLQ 535
          ||||||||| || ||||:|||||||:|||||||| |||:|:||:| | | || || ||
Db    710 AVCRDLSEALKAVPSLRELGLLQNRLTEAGLRLLSQGLAWPCKVQTLRICMPGLQEVIH 769

Qy    536 YLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHCGCGLQTLSLASVELSEQSLQELQ 595
          ||:::| |:|| |||||:|| || |:|::| :|:|:|| |:|| :|:|: ||: :||:||
Db    770 YLVILVQQSPVLTTLDLSGCCLPGTVVEPLCSALKHPKCGJXTLSLTSVELTENPLRELQ 829

Qy    596 AVRRAKPDLVITHPALDGHPQPPK 619
          ||:| ||| |:||| |:||:||| |
Db    830 AVKTLKPDLAIIHSKLGTHPQPLK 853

RESULT 3
PYA5_MOUSE
ID   PYA5_MOUSE     STANDARD;     PRT;     843 AA.
AC   Q91WS2; Q8K0L4;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   PYRIN-containing APAF1-like protein 5-like.
GN   PYPAF5.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   CONCEPTUAL TRANSLATION OF 1-287.
RA   Hinz U.;
RL   Unpublished observations (FEB-2003).
RN   [2]
RP   SEQUENCE OF 288-843 FROM N.A.
RC   TISSUE=Kidney;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,

RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marsina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
RN   [3]
RP   IDENTIFICATION OF MAMMALIAN ORTHOLOGS OF PYPAF5.
RA   Albrecht M., Domingues F.S., Schreiber S., Lengauer T.;
RT   "Identification of mammalian orthologs associates PYPAF5 with distinct
RT   functional roles.";
RL   FEBS Lett. 538:173-177(2003).
CC   -!- FUNCTION: May mediate activation of CASP1 via ASC and promote
CC       activation of NF-kappa-B (By similarity).
CC   -!- SUBUNIT: Binds to ASC with its DAPIN domain. (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 3 leucine-rich (LRR) repeats.
CC   -!- CAUTION: The N-terminus was extended using ESTs and genomic
CC       sequences, in analogy to ortholog sequences.
DR   EMBL; BC013519; AAH13519.1; -.
DR   EMBL; BC031139; AAH31139.1; ALT_INIT.
DR   MGD; MGI:2141990; Pypaf5.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RN:nh.
DR   Pfam; PF00560; LRR; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat.
FT   DOMAIN       11    102       DAPIN.
FT   DOMAIN      168    485       NACHT.
FT   DOMAIN      569    585       ASP/GLU-RICH.
FT   DOMAIN      654    661       POLY-LYS.
FT   REPEAT      434    459       LRR 1.
FT   REPEAT      609    632       LRR 2.
FT   REPEAT      811    834       LRR 3.
FT   NP_BIND     174    181       ATP (POTENTIAL).
SQ   SEQUENCE   843 AA;  94592 MW;  35FB7A766A47DB51 CRC64;

Query Match                  66.1%;  Score 2127.5;  DB 1;  Length 843;
Best Local Similarity        70.2%;          Pred. No. 9.6e-148;
Matches 431;  Conservative  63; Mismatches 103; Indels 17; Gaps 7;

Qy      1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTT 60
          :||| |||||||||||||||| ||||||||||||:| :|||||||| ||:|| ||||||||
Db    239 ILAQPNRLLFILDGADELPLPSSEATPCKDPLEATSGLRVLSGLLSQELLPPGARLLVTT 298

Qy     61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCF 120
          ||| |||||||||||||||||||||||||||||||||:|||||||||||||||||||||||
Db    299 RHAATGRLQGRLCSPQCAEIRGFSDKDKKKYFFKFFRDERKAERAYRFVKENETLFALCF 358

```
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
DR   EMBL; AF466632; AAL90874.1; -.
DR   MGD; MGI:2653833; Cias1.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   SMART; SM00368; LRR_RI; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Apoptosis; Repeat; Leucine-rich repeat.
FT   DOMAIN       1     91       DAPIN.
FT   DOMAIN     216    532       NACHT.
FT   REPEAT     737    760       LRR 1.
FT   REPEAT     794    817       LRR 2.
FT   REPEAT     851    874       LRR 3.
FT   REPEAT     880    903       LRR 4.
FT   REPEAT     908    931       LRR 5.
FT   REPEAT     937    964       LRR 6.
FT   REPEAT     965    988       LRR 7.
SQ   SEQUENCE  1033 AA;  118274 MW;  59246909668212117 CRC64;

Query Match
Best Local Similarity   25.9%;   Score 835;  DB 1;  Length 1033;
                        Pred. No. 3.1e-53;
Matches  226;  Conservative  112;  Mismatches  239;  Indels  134;  Gaps  15;

Qy    1 MLAQPQRLLFILDGADELP-ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLVT   59
          : :|| |||||| ||::  |||||||||::|::||: |:|:|
Db  286 ILRKPSRILFLMDGFDELQGAFDEHIGEVCTDWQKAVRGDILLSSLIRKLLDKASLLIT  345

Qy   60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFPRDERRAERAYRFVKENETLFALC  119
         |||::|::|:|:: ||:||:|||::|| :::PHSHJVLTTRFLFGLLSA
Db  346 TRPVALEKLQHLLDHPRHVEILGFSEAKRKEYFFKYFSNELQAREAFRLIQENEVLFTMC  405

Qy  120 FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNL  179
         |||||||||||:||:||||:|||:|||||||||:||:::||||:||::||:||||||:
Db  406 FIPLVCWIVCTGLKQQMETGKSLAQTSKTTTAVYVFFLSSLLQSRGGIEEHLFSDYLQGL  465

Qy  180 CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF  239
         | :||:|:| |||: |::::|::|::|::| ||||::|:|||:||:|||||||||:
Db  466 CSLAADGIWNQKILFEECDLRKHGLQKTDV-SAFLRMNVFQKEVDCERFYSFSHMTFQEF  524

Qy  240 LAALSYLLE--DGGVPRTAAGGVGTLLRGDAQ---------PHSHJVLTTRFLFGLLSA  287
         |||||||    | || | ||| |::  ::| |          || |||::||::
Db  525 FAAMYLLEEEAEGETVRKGPGGCSDLLNRDVKVLLENYGKFEKGY:FVVRFLFG:VNQ  584

Qy  288 ERMRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEBPEEEEGE  347
         |:  |: :|| | : : | | | ::||||
Db  585 ERTSYLEKKLSCKISQQVRLELLKWI------------EVKAKAKLQ--WQPSQ----  625

Qy  348 EPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALR  407
         : ||::|||||||||| ::||||| | || |||: |: : : :||| :|||||:
Db  626 ----LELFYCLYEMQEEDFVQSAMDHFPKIEIN--LSTRMDHVVSSFCIKNCHRVKTLS  678

Qy  408 LISCRLVAAQEKKKKSLGKRLQ--------------ASLGGG-SSQGTTK  442
         |: |:|  |: : |: || |                |||| |:|| :
Db  679 LGFFHNSPKEEEEERRGGRPLDQVQCVFPDTHVACSSRLVNCCLTSSPCRGLFSSLSTNR  738

Qy  443 QL---------------PASLLHFLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPA  490
          :               |::||:||||::|  ::| :| |||:
Db  739 SLTELDLSDNTLGDFG--MRVLCEALQHPGCNIQRLWLGRCGLSHQCCFDISSVLSSSQK  796

Qy  491 LTEGLLHNRLSEAGLRMLSEGLAWPQCR----------------------  519
         |:| |: |:|| :|| :|
Db  797 LVELDLSDNALGDFGIRLLCVGLKGHLLCNLQKLWLVSCCLTSACCQDLALVLSSNHSLTR  856

Qy  520 ------------------VQTVRVQLPDPQRGLQYL-----  537
                            | ||| |||
```

```
Qy  121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNLC  180
         |||||||||||:||||||||||||||||||||||||||:||:: |:|||:||:||:||:|
Db  359 VPFVCWIVCTVLQQQLELGRDLSRTSKTTTSVYLLFITSMLKSAG-TNGPRVQGELRTLC  417

Qy  181 RLAREGVLG-RRAQFAEKELEQLELRGSKVQTLF-SKKELPGVLETEVTYQFIDQSFQEF  239
         ||||||:|   :|||::|:|:|:|||||:|||:| :||::|||:|:|||||||||||:|||
Db  418 RLAREGILDHHKAQFSEEDLEKLKLRGSQVQTIFLNKKEIPGVLKTEVTYQFIDQSPQEF  477

Qy  240 LAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHZVLTRFLFGLLSAEXMRDIERHFGC  299
         |||||||||:||:||||||||  :|::|||| ::::||||||||:: :||||||:|||
Db  478 LAALSYLLEAERTPGTPAGGVQKLLNSDAELRGHLAZTRFLFGLLNEGLRDIGNHFGC  537

Qy  300 MVSERVKQEALRWVQGQG--QGCPGVAPEVTEGAKGLEDTEEPEEEGEEF-NYPLELL  356
         :|:::|:|: |||||||   :||| |||||  |   ||::|::||||: |  ||||||
Db  538 VVPDHVKKQTLRWVQGQSHPKGPPVGAKKTAE---LEDIEDAEEEEEDLNFGLELL  593

Qy  357 YCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLSCRLVAA  416
         ||||||||||:||||::|:||||||||||||||||||||||||||||:|||::||::|||
Db  594 YCLYETQEDFVRQALSSLPEIVLERVRLTRMDLEVLRYCVQCCPDGQALRLVSCGLVAA  653

Qy  417 QE--KKKKSLGKRLQASLGGGSSQGTTKGLNRLSEAGLRMLSEGLAWPQCRVRVQLPDPQRGL  474
         ||  ||:::|:: |:                :|||:::||||:|:||:||||||:||||:|
Db  654 KEKKKKKKSLVKRLK------GSQSTKKQPPVSLLRPLCETMTTPKCHLSVLLSHCRLP  707

Qy  475 DAVCRDLSEALRAAPALTELGLLNRLSEAGLRMLSEGLAWPQCRVCTVRVQLPDPQRGL  534
         ||||||||||:|||||||:||||||:|:: |:||||:|:|||||||||:||||||:||:|
Db  708 DAVCRDLSEALKVAPAREGLLQGRLTNTGLRLLCEGLAWPKCQVKTLRMQLPDLQEVI  767

Qy  535 QYLVGM:RGSPALTTLDLSGCQLZAPMVTYLCAV:QHQCCGIQTLSLASVELSECSLQEL  594
         ||||||:|||||||||||||||:|||||||::|||:|:| | |:|:|::||||:||:|:|:
Db  768 NYLVIV:QQSPVLTTLDLSGCQLPGVIVEPLCAALHKPKCSLKTLSLTSVELSENSLRDL  827

Qy  595 QAVKRAKPDLVTH  608
         ||||:||||:|||
Db  828 QAVTSKPDLSIIY  841
```

```
RESULT 4
CIS1_MOUSE
ID   CIS1_MOUSE   STANDARD;   PRT;   1033 AA.
AC   Q8R4B8;
DT   15-SEP-2003 (Rel. 42, Created)
DT   15-SEP-2003 (Rel. 42, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cold autoinflammatory syndrome 1 protein homolog (PYRIN-containing
DE   APAF1-like protein 1) (Mast cell maturation inducible protein 1).
GN   CIAS1 OR PYPAF1 OR MMIG1
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=BALB/cJ;
RA   Kikuchi-Yanoshita R., Koga K., Taketomi Y., Sugiki T., Saito T.,
RA   Ishii S., Hisada M., Suzuki-Nishimura T., Uchida M.K., Moon T.-C.,
RA   Chang H.-W., Sawada M., Inagaki N., Nagai N., Murakami M., Kudo I.;
RT   "Identification of inducible genes during in vitro maturation of mouse
RT   bone marrow-derived mast cells to connective tissue-type mast cells.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: May function as a potential inducer of apoptosis.
CC       Interacts selectively with apoptosis-associated specklike protein
CC       containing a CARD domain (ASC). This complex may function as an
CC       upstream activator of NF-kappaB signaling (By similarity).
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 7 leucine-rich (LRR) repeats.
CC   ----------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
```

```
Db    857  LYIGRNALGDSGVQVLCERMKDPQCNLQKLGLVNSGLTSICCSALTSVLKTNQNFTHEYL 916

Qy    538  ------------VGMLRQSPALT-LDLSGCQLPAP------------------------- 560

Db    917  RSNALGDTGLRLLCEGLLHPDCKLQMLELDNCSLTSHSCWNLSTILTHNHSLRKLNLGNN 976

Qy    561  ------MVTYLCAVLQHQGGGCQTLSLASVELSEQSLQELQAVKRAKPDLVI 606
               ||| ::| ||:| ::| |:  |||
Db    977  DLGDLCVVLCEVLKQQGGCLLQSLQLGEMYLNETKRALEALQEEKPELTI 1027
```

RESULT 5
ID  PYA7_HUMAN     STANDARD;      PRT;   1062 AA.
AC  P59046;
DT  28-FEB-2003 (Rel. 41, Created)
DT  28-FEB-2003 (Rel. 41, Last sequence update)
DT  15-SEP-2003 (Rel. 42, Last annotation update)
DE  PYRIN-containing APAF1-like protein 7 (Monarch-1).
GN  PYPAF7.
OS  Homo sapiens (Human).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX  NCBI_TaxID=9606;
RN  [1].
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RX  MEDLINE=22162427; PubMed=12019269;
RA  Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA  Lora J.M., Geddes B.C., Briskin M., DiStefano P.S., Bertin J.;
RT  "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT  activation of NF-kappa B and caspase-1-dependent cytokine
RT  processing.";
RL  J. Biol. Chem. 277:29874-29880(2002).
RN  [2]
RP  SEQUENCE FROM N.A. (ISOFORMS 1; 2 AND 3).
RC  TISSUE=Lymphoma;
RX  MEDLINE=22388257; PubMed=12477932;
RA  Williams K.L., Linhoff M.W., Harton J.A., Ting J.P.Y.;
RL  Submitted (MAY-2002) to the EMBL/GenBank/DDBJ databases.
RN  [3]
RP  SEQUENCE FROM N.A. (ISOFORM 1).
RC  TISSUE=Leukocyte;
RX  MEDLINE=22388257; PubMed=12477932;
RA  Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA  Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA  Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA  Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA  Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA  Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA  Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA  Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA  Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA  Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA  Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA  Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA  Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA  Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA  Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA  Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA  Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT  "Generation and initial analysis of more than 15,000 full-length
RT  human and mouse cDNA sequences.";
RL  Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002)
CC  -!- FUNCTION: May mediate activation of CASP1 via ASC and promote
CC      activation of NF-kappa-B via IKK.
CC  -!- SUBUNIT: Binds to ASC with its DAPIN domain.
CC  -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC  -!- ALTERNATIVE PRODUCTS:
CC      Event=Alternative splicing; Named isoforms=3;
CC      Name=1; Synonyms=I;
CC        IsoId=P59046-1; Sequence=Displayed;
CC      Name=2; Synonyms=II;
CC        IsoId=P59046-2; Sequence=VSP_005524;
CC      Name=3; Synonyms=III;
CC        IsoId=P59046-3; Sequence=VSP_005523;
CC  -!- TISSUE SPECIFICITY: Detected only in peripheral blood leukocytes,
CC      predominantly in eosinophils and granulocytes, and at lower levels
CC      in monocytes.
CC  -!- SIMILARITY: Contains 1 DAPIN domain.
CC  -!- SIMILARITY: Contains 1 NACHT domain.
CC  -!- SIMILARITY: Contains 8 leucine-rich (LRR) repeats.

CC  -------------------------------------------------------------------------
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -------------------------------------------------------------------------
DR  EMBL; AY095146; AAM18227.1; -.
DR  EMBL; AY116204; AAM75142.1; -.
DR  EMBL; AY116205; AAM75143.1; -.
DR  EMBL; AY116206; AAM75144.1; -.
DR  EMBL; BC028069; AAH28069.1; -.
DR  InterPro; IPR001611; LRR.
DR  InterPro; IPR007091; LRR_RNinh.
DR  InterPro; IPR003590; LRR_RNinh_sub.
DR  InterPro; IPR007111; NACHT_NTPase.
DR  InterPro; IPR004020; PAAD_DAPIN_dom.
DR  Pfam; PF00560; LRR; 2.
DR  Pfam; PF02758; PAAD_DAPIN; 1.
DR  PRINTS; PR00019; LEURICHRPT.
DR  SMART; SM00368; LRR_RI; 11.
DR  PROSITE; PS50824; DAPIN; 1.
DR  PROSITE; PS50837; NACHT; -.
KW  ATP-binding; Leucine-rich repeat; Repeat; Alternative splicing.
FT  DOMAIN       1     95       DAPIN.
FT  DOMAIN     211    528       NACHT.
FT  REPEAT     713    736       LRR 1.
FT  REPEAT     742    765       LRR 2.
FT  REPEAT     770    792       LRR 3.
FT  REPEAT     799    822       LRR 4.
FT  REPEAT     827    850       LRR 5.
FT  REPEAT     884    907       LRR 6.
FT  REPEAT     941    969       LRR 7.
FT  REPEAT     998   1021       LRR 8.
FT  NP_BIND    217    224       ATP (POTENTIAL).
FT  VARSPLIC   863    974       Missing (in isoform 3).
FT                              /FTId=VSP_005523.
FT  VARSPLIC   977   1032       Missing (in isoform 2).
FT                              /FTId=VSP_005524.
FT  CONFLICT   692    692       MISSING (IN REF. 1 AND 3).
SQ  SEQUENCE   1062 AA;  120328 MW;  0AB81C87F16497F CRC64;

Query Match             25.6%;  Score 825;  DB 1;  Length 1062;
Best Local Similarity   32.6%;  Pred. No. 1.7e-52;
Matches  219;  Conservative 104;  Mismatches 245;  Indels 104;  Gaps  15;

Qy     1  MLAQPQRLLFILDGADEL-PALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 59
          ||| |  |::|:|| ::|:   :|||: :|:|  | | ::|:||  ||:|: |: ||||| :
Db   282  LIRVPERLLFILIDGFDELKPSFHDPOGPWCLCWEEKRPTELLLNSLIRKKLLPELSLLT 341

Qy    60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRABRAYRFVKENETLFALC 119
          ||||| | :| | :| |   |  |:  : ||||:| ||  | |  |||||:|: |: |||
Db   342  TRPTALEKLHRLLEHPRHVEILGFSEAERKEYFYKYFHNAEQAGQVFNVVRDNEPLFTMC 401

Qy   120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFITSVLSSAPVADGPRLQ--GDLR 177
          |||||  ||  :|:| |::|::: : |:::: ||:|::|| :|| :::|:  :::   :|||
Db   402  FVPLVCWVVCTLCQQLELGGGLLRQTSRTTTAVMLYLLSLMQPKGA--PRLQPPNQR 459

Qy   178  NLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLPLSKKELPGVLETEVTYQFIDQSFQ 237
          ||| ::|| | | :|| :| ||::|  | :: : ::::: :  :|: |:| ::|:   ||:|
Db   460  GLCSLAADGLWNQKILFEEQDLRKHGLDGEDV-SAFLNMNIFQKDINCERYYSFIHLSFQ 518

Qy   238  EFLAALSYLLEDGGVPRTAGGVG---------TLLRGDAQPHSHLVLTRFLGLLSAERM 290
          |||  :||::||  ||  |   |            ||| |  |:|| ||:| || || |
```

```
Db   519  EFFAAMYYILDEG------EGGAGPDQDVTRLLTEYAFSERSFLALTSRFLFGLLNEETR  572
          |    :::|  |:| :           ||:|:|:::|:
Qy   291  RDIERHFGCMVSERVKQEALRWVQQGQGCPGVAPEVTEGAKG-EDTEEPEEEGEEPN    350

Db   573  SHLEKSLCWKVSPHIKMDLLQWIQSKAQ------------SDGSTLQQGS---       610
Qy   351  YPDELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLS  410
          :     ||| |: |: |: |:|     |   |  |:::|     |: |
Db   611  --LEFFSCLYEIQEEFIQQALSHQVIVVSNIA-SKMEHMVSSFCLKRCRSAQVLRLYG   667

Qy   411  CRLVAAQEKKKKSLGKRLQASLGGGS--------SQGTTKQLPASL--------       446
          :  |   :                 |
Db   668  ATYSADGE------DRARCSAGAHTLLVQLRPERTVLLCAYSEHLAAALCTNPNLIELS  720

Qy   449  ------LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGJLH    495
                :|:|:     |:  | : |  ||  |   ||:|:||  |||:| :|
Db   721  LYRNA-GSRGVK-LCQGLRHPNCKLQNLRLKRCRISSSACEDLSAALIANKNLTRMDLSG 780

Qy   499  NRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRG-LQYLVGMLRQSPALTTLDLSGCQL  557
          |  |  ||  |||| ||:|:| |  :|:|    :  |  |  | : |  |:|| |
Db   781  NGVGFPGMMLLCEGLRHPQCRLQMQLRKCQLESGACQEMASVLGTNPHLVELDLTGNKL  840

Qy   558  PAPWVTYLCAVLQHQGCGLGTL----------SLASVELSEQSLQELCAVRRAKPEL   604
                  | | | |:| ||| |               |: ||: | |||  |
Db   841  EDLGLRLLCQGLRHPVCRLRTFWLKICRJTAAACDELASTLSVNQSLRELDLSLNELGDL 900

Qy   605  VI------THP                                                  609
Db   901  GVLLCEGLRHP                                                  912
```

RESULT 6
CIS1_HUMAN
ID   CIS1_HUMAN   STANDARD;   PRT;   1034 AA.
AC   Q96P20; C75434; Q8TCW0; Q8TEC9; Q8WXH9;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   Cold autoinflammatory syndrome 1 protein (Cryopyrin) (NACHT-, LRR- and
DE   PYD-containing protein 3) (PYRIN-containing protein 1;
DE   (Angiotensin/vasopressin receptor AII/AVP-like).
GN   CIAS1 OR NALP3 OR PYPAF1
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2), AND VARIANT FCAS MET-198; VAL-439
RP   AND GLY-627, AND VARIANT MWS VAL-352.
RX   MEDLINE=21547523; PubMed=11687797;
RA   Hoffman H.M., Mueller J.L., Broide D.H., Wanderer A.A.,
RA   Kolodner R.D.;
RT   "Mutation of a new gene encoding a putative pyrin-like protein causes
RT   familial cold autoinflammatory syndrome and Muckle-Wells syndrome.";
RL   Nat. Genet. 29:301-305(2001).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=21969508; PubMed=11786556;
RA   Manji G.A., Wang L., Geddes B.J., Brown M., Merriam S., Al-Garawi A.,
RA   Mak S., Lora J.M., Briskin M., Jurman M., Cao J., DiStefano P.S.,
RA   Bertin J.;
RT   "PYPAF1, a PYRIN-containing APAF1-like protein that assembles with ASC
RT   and activates NF-kB.";
RL   J. Biol. Chem. 277:11570-11575(2002).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1, 2 AND 3), VARIANTS FCAS MET-198, AND
RP   VARIANTS FCAS/MWS TRP-260 AND PRO-305.
RX   MEDLINE=22241234; PubMed=12355493;
RA   Aganna E., Martinon F., Hawkins P.N., Ross J.B., Swan D.C.,
RA   Booth D.R., Lachmann H.J., Gadet R., Woo P., Feighery C.,
RA   Cotter F.E., Thome M., Hitman G.A., Tschopp J., McDermott M.F.;
RT   "Association of mutations in the NALP3/CIAS1/PYPAF1 gene with a broad
RT   phenotype including recurrent fever, cold sensitivity, sensorineural
RT   deafness, and AA amyloidosis.";
RL   Arthritis Rheum. 46:2445-2452(2002).
RN   [4]
RP   SEQUENCE OF 391-1034 FROM N.A. (ISOFORM 1).
RC   TISSUE=Blood;
RX   MEDLINE=20499367; PubMed=11042152;
RA   Zhang Q.-H., Ye M., Wu X.Y., Ren S.X., Zhao M., Zhao C.J., Fu G.,
RA   Shen Y., Fan H.Y., Lu G., Zhong M., Xu X.R., Han Z.-G., Zhang J.W.,
RA   Tao J., Huang Q.-H., Zhou J., Hu G.X., Gu J., Chen S.-J., Chen Z.;
RT   "Cloning and functional analysis of cDNAs with open reading frames for
RT   300 previously undefined genes expressed in CD34+ hematopoietic
RT   stem/progenitor cells.";
RL   Genome Res. 10:1546-1560(2000).
RN   [5]
RP   VARIANT FCAS MET-198, VARIANTS MWS ASN-303; MET-348; THR-439 AND
RP   ARG-569, AND VARIANT FCAS/MWS TRP-260.
RX   MEDLINE=21987640; PubMed=11992256;
RA   Dode C., Le Du N., Cuisset L., Letourneur F., Berthelot J.-M.,
RA   Vaudour G., Meyrier A., Watts R.A., Scott D.G.I., Nicholls A.,
RA   Granel B., Frances C., Garcier F., Edery P., Boulinguez S.,
RA   Domergues J.-P., Delpech M., Grateau G.;
RT   "New mutations of CIAS1 that are responsible for Muckle-Wells syndrome
RT   and familial cold urticaria: a novel mutation underlies both
RT   syndromes.";
RL   Am. J. Hum. Genet. 70:1498-1506(2002).
RN   [6]
RP   VARIANTS CINCA ASN-303; SER-309; ARG-358; ASN-436; SER-573 AND
RP   THR-662, AND TISSUE SPECIFICITY.
RX   MEDLINE=22062556; PubMed=12032915;
RA   Feldmann J., Prieur A.-M., Quartier P., Berquin P., Certain S.,
RA   Cortis E., Teillac-Hamel D., Fischer A., de Saint Basile G.;
RT   "Chronic infantile neurological cutaneous and articular syndrome is
RT   caused by mutations in CIAS1, a gene highly expressed in
RT   polymorphonuclear cells and chondrocytes.";
RL   Am. J. Hum. Genet. 71:198-203(2002).
CC   -!- FUNCTION: May function as a potential inducer of apoptosis.
CC       Interacts selectively with apoptosis-associated specklike protein
CC       containing a CARD domain (ASC). This complex may function as an
CC       upstream activator of NF-kappaB signaling.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC       Name=2;
CC         IsoId=Q96P20-1; Sequence=Displayed;
CC       Name=1;
CC         IsoId=Q96P20-2; Sequence=VSP_005520, VSP_005521;
CC       Name=3;
CC         IsoId=Q96P20-3; Sequence=VSP_005519;
CC   -!- TISSUE SPECIFICITY: Expressed in blood leukocytes. Strongly
CC       expressed in polymorphonuclear cells, undetectable or expressed
CC       at a lower magnitude in B and T lymphoblasts, respectively. High
CC       level of expression detected in chondrocytes. Low or no expression
CC       in the other tissues tested.
CC   -!- DISEASE: Defects in CIAS1 are a cause of familial cold
CC       autoinflammatory syndrome (FCAS), commonly known as familial cold
CC       urticaria. FCAS is rare autosomal dominant systemic inflammatory
CC       disease characterized by episodes of rash, arthralgia, fever and
CC       conjunctivitis after generalized exposure to cold.
CC   -!- DISEASE: Defects in CIAS1 are a cause of Muckle-Wells syndrome
CC       (MWS), a rare autosomal dominant fever syndrome with episodic
CC       urticaria, arthralgia, amyloidosis and progressive sensorineural
CC       deafness.
CC   -!- DISEASE: Defects in CIAS1 are the cause of chronic infantile
CC       neurologic cutaneous and articular syndrome (CINCA), also known as
CC       'neonatal onset multisystem inflammatory disease,' or NOMID, a
CC       rare congenital inflammatory disorder characterized by a triad of
CC       neonatal onset of cutaneous symptoms, chronic meningitis, and
CC       joint manifestations with recurrent fever and inflammation.
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 7 leucine-rich (LRR) repeats.
CC   -!- CAUTION: Ref.4 sequence differs from that shown due to frameshifts
CC       in positions 893, 918 and 926.

```
DR   EMBL; AF410477; AAL33908.1; -.
DR   EMBL; AF427617; AAL33911.1; -.
DR   EMBL; AY051117; AAL12497.1; JOINED.
DR   EMBL; AY051112; AAL12497.1; JOINED.
DR   EMBL; AY051113; AAL12497.1; JOINED.
DR   EMBL; AY051114; AAL12497.1; JOINED.
DR   EMBL; AY051115; AAL12497.1; JOINED.
DR   EMBL; AY051116; AAL12497.1; JOINED.
DR   EMBL; AY056059; AAL12497.1; JOINED.
DR   EMBL; AY056060; AAL12497.1; JOINED.
DR   EMBL; AY051117; AAL12498.1; -.
DR   EMBL; AY051112; AAL12498.1; JOINED.
DR   EMBL; AY051113; AAL12498.1; JOINED.
DR   EMBL; AY051114; AAL12498.1; JOINED.
DR   EMBL; AY051115; AAL12498.1; JOINED.
DR   EMBL; AY051116; AAL12498.1; JOINED.
DR   EMBL; AF420469; AAL65136.1; -.
DR   EMBL; AF468522; AAL78632.1; ALT_INIT.
DR   EMBL; AY092033; AAM14669.1; ALT_INIT.
DR   EMBL; AF418985; AAL64640.2; ALT_INIT.
DR   EMBL; AF054176; AAC39910.1; ALT_FRAME.
DR   Genew; HGNC:16400; CIAS1.
DR   MIM; 606416; -.
DR   MIM; 120100; -.
DR   MIM; 191900; -.
DR   MIM; 607115; -.
DR   GO; GO:0016506; F:apoptosis activator activity; NAS.
DR   GO; GO:0006917; P:induction of apoptosis; NAS.
DR   GO; GO:0006954; P:inflammatory response; IMP.
DR   GO; GO:0007165; P:signal transduction; NAS.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR003590; LRR_RNinh_sub.
DR   InterPro; IPR007111; NACH NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF00560; LRR; 2.
DR   SMART; SM00368; LRR_RI; 3.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Apoptosis; Repeat; Leucine-rich repeat; Alternative splicing;
KW   Disease mutation; Deafness.
FT   DOMAIN      218       91 DAPIN.
FT   DOMAIN      534          NACHT.
FT   REPEAT      738      761 LRR 1.
FT   REPEAT      795      818 LRR 2.
FT   REPEAT      852      875 LRR 3.
FT   REPEAT      881      904 LRR 4.
FT   REPEAT      909      937 LRR 5.
FT   REPEAT      939      961 LRR 6.
FT   REPEAT      966      989 LRR 7.
FT   DOMAIN      688     1034 POLY-GLU.
FT   VARSPLIC    718      695 Missing (in isoform 3).
FT                            /FTId=VSP_005519.
FT   VARSPLIC    719      775 Missing (in isoform 1).
FT                            /FTId=VSP_005520.
FT   VARSPLIC    834      890 Missing (in isoform 1).
FT                            /FTId=VSP_005521.
FT   VARIANT     198      198 V -> M (IN FCAS AND MWS).
FT                            /FTId=VAR_013227.
FT   VARIANT     260      260 R -> W (IN FCAS AND MWS).
FT                            /FTId=VAR_014104.
FT   VARIANT     303      303 D -> N (IN CINCA AND MWS).
FT                            /FTId=VAR_014105.
FT   VARIANT     305      305 L -> P (IN FCAS AND MWS).
FT                            /FTId=VAR_014124.
FT   VARIANT     309      309 F -> S (IN CINCA).
FT                            /FTId=VAR_014106.
FT   VARIANT     348      348 T -> M (IN MWS).
FT                            /FTId=VAR_014366.
FT   VARIANT     352      352 A -> V (IN MWS).
FT                            /FTId=VAR_013228.
FT   VARIANT     358      358 H -> R (IN CINCA).
FT                            /FTId=VAR_014367.
FT   VARIANT     436      436 T -> N (IN CINCA).
FT                            /FTId=VAR_014368.
FT   VARIANT     439      439 A -> T (IN MWS).
FT                            /FTId=VAR_014369.
FT   VARIANT     439      439 A -> V (IN FCAS).
FT                            /FTId=VAR_013229.
```

```
Query Match:             24.9%;   Score 800.5;   DB 1;   Length 1034;
Best Local Similarity   28.4%;    Pred. No. 1e-50;
Matches 219;   Conservative 117;   Mismatches 241;   Indels 193;   Gaps 17;

Qy    1 MLAQPQRLFILDGADELPALGGPEAAP-CTDPFEAASGARVLGGLLSKALLPTALLVT 59
          :.:|:|||:.::||| ||                :|||:|:.|: |::
Db  288 IVRKPSRILFLMDGFDELQGAFDEHIGPLCTDWQKAERGDILLSSLIRKKLLPEASLLIT 347

Qy   60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
         || :.|: | |:   |.|:.|.|||:. :|::||:|| |:|  ||.|:|::|||:|.:|
Db  348 TRPVALEKLQHLLDHPRHVEILGFSEAKREYFFKYFSDEAQARAAFSLIQENEVLFTMC 407

Qy  120 FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVLLFITSVLSSAPVADGPRLQGDLRNL 179
         |.|.||||||| |:||:| |..|::|||||:.|::|:.|:|::.     |  :|  .:|
Db  408 FIPLVCWIVCTGLKQQMESGKSLAQTSKTTAVVVFFLSSLLQPRGG-SQEHGLCAHLWGL 467

Qy  180 CRLAREGVLGRRAQFAEKELEQLBLRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF 239
         | || :|: :.:: :.|::|  :   .  .: |  ::|: ..:|.  |..|:  :.:FQEF
Db  468 CSLAADGIWNQKILFEESDLRNHGLQKADV-SAFLRMNLFQKEVDCEKFYSFIHMTFQEF 526

Qy  240 LAALSYLLEDGGVPRTAAGG----------------VGTLLRGDAQ-PHSHLVLTRFLFGLLSAER 289
         .|| .|||:::. |                         :  |    |.  ::: ||||||:.:||
Db  527 FAAMYLLEEEKEGRTNVPGSRLKLPSRDVTVLENYGKFPEKGYLIFVVRFLFGLVNQER 586

Qy  290 MRDIERHFGCMVSERVKQEALRWVGQGQGCPGVAPEVTEGAKGLED*EEPEEEEGEEP 349
          .|:.    |.:.|::  :|:| |                EVKAKAKLQ--:QPSQ-
Db  587 TSYLEKKLSCKISQQIRLELLKWI------------EVKAKAKLQ--:QPSQ- 625

Qy  350 NYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLI 409
          :..|||||:||::|::.|.|:|         |:.  .|::|::.| ||.|:| ..|  :.
Db  626 ---LELFYCLYEMQEEDFVQRAMDYFPKIEIN--LSTRMDHMVSSFCIENCHRVESLSLG 680

Qy  410 SCRLVAAQEKKKKSLGKRLQ----------ASLGGGSQG-TTKQLPASLLHPLF------QA 455
          .  |:  :|:::  |                :|.  .  :.  |:  |: |: |:
Db  681 FLHNMPKEEEEEKEGRH.DMVQCVLPSSSHAACSHGLVNSHLTSSFCRGLFSVLSTSQS 740

Qy  456 VT------------DP--- 459
         :.            :
Db  741 LTELDLSDNSLGDPGMRVLCETLQHPGCNIRRLWLGRCGLSHECCFDISLVLSSNQKLVE 800

Qy  460 ---------------LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGL 460
                        |||.||:|.|:||||:|.|.||:|::   .| :|  |.
Db  801 LDLSDNALGDFGIRLLCVGLKGHLLCNLKWLVSCCLTSACCQDLASVLSTSHSLTRLYV 860

Qy  497 LHNRLSEAGLRMLSEGLAWPQCRVQTVRV------- 525
         |.| | ::|:::|||||..||| |||:.
Db  861 GENALGDSGVAILCEKAKNPQCNLQKLGLVNSGLTSVCCSALSVLSTNQNLTHLYLRGN 920

Qy  526 QLPDPQRGLQYLV-GMLRQSPALTTLDLSGCQLPAPM--- 561
         .|.|  .|:  :| |:|:|. ||..|:|  |.|  :
Db  921 TLGD--KGIKLLCEGLLHPDCKLQVLEDNCNLTSHCCWDLSTLLTSSQSLRKLSLGNND 978

Qy  562 ----VTYLCAVLQHQGCLQOTLSLASVELSEQSLQELQAVKRAKPDLVI 606
             :||| .||.::.|| |.||:|::|||:||||:||.
Db  979 LGDLGVMMFCEVLKQQSCLLQNLGLSEMYPNYETKSALETLQEEKPELTV 1028
```

RESULT 7

```
NALI_HUMAN
ID   NAL1_HUMAN    STANDARD;      PRT;  1473 AA.
AC   Q9C000; Q9BZZ8; Q9BZZ9; Q9HAV8; Q9UFT4; Q9Y2E0;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   NACHT-, LRR- and PYD-containing protein 2 (Death effector filament-
DE   forming ced-4-like apoptosis protein) (Nucleotide-binding domain and
DE   caspase recruitment domain) (Caspase recruitment domain protein 7).
GN   NALP1 OR DEFCAP OR NAC OR CARD7 OR KIAA0926.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=21169419; PubMed=11270363;
RA   Bertin J., DiStefano P.S.;
RT   "The PYRIN domain: a novel motif found in apoptosis and inflammation
RT   proteins.";
RL   Cell Death Differ. 7:1273-1274(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=21148092; PubMed=11250163;
RA   Martinon F., Hofmann K., Tschopp J.;
RT   "The pyrin domain: a possible member of the death domain-fold family
RT   implicated in apoptosis and inflammation.";
RL   Curr. Biol. 11:R118-R120(2001).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORMS 1 AND 2).
RC   TISSUE=Erythro-leukemia;
RX   MEDLINE=21153743; PubMed=11076957;
RA   Hlaing T., Guo R.-F., Dilley K.A., Loussia J.M., Morrish T.A.,
RA   Shi M.M., Vincenz C., Ward P.A.;
RT   "Molecular cloning and characterization of DEFCAP-L and -S, two
RT   isoforms of a novel member of the mammalian Ced-4 family of apoptosis
RT   proteins.";
RL   J. Biol. Chem. 276:9230-9238(2001).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORMS 1; 2; 3 AND 4), AND PROTEIN INTERACTION.
RC   TISSUE=T-cell;
RX   MEDLINE=21153744; PubMed=11113115;
RA   Chu Z.-L., Pio F., Xie Z., Welsh K., Krajewska M., Krajewski S.,
RA   Godzik A., Reed J.C.;
RT   "A novel enhancer of the Apaf1 apoptosome involved in cytochrome
RT   c-dependent caspase activation and apoptosis.";
RL   J. Biol. Chem. 276:9239-9245(2001).
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RC   TISSUE=Brain;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Nagase T., Ishikawa K.-I., Suyama M., Kikuno R., Hirosawa M.,
RA   Miyajima N., Tanaka A., Kotani H., Nomura N., Ohara O.;
RT   "Prediction of the coding sequences of unidentified human genes. XII.
RT   The complete sequences of 100 new cDNA clones from brain which code
RT   for large proteins in vitro.";
RL   DNA Res. 6:63-70(1999).
RN   [6]
RP   SEQUENCE OF 282-1473 FROM N.A. (ISOFORM 1).
RC   TISSUE=Uterus;
RX   MEDLINE=99246063; PubMed=10231032;
RA   Koehrer K., Beyer A., Mewes H.-W., Gassenhuber J., Wiemann S.;
RL   Submitted (SEP-1999) to the EMBL/GenBank/DDBJ databases.
CC   -!- FUNCTION: Able to form cytoplasmic structures termed death
CC       effector filaments. Enhances APAF1 and cytochrome c-dependent
CC       activation of pro-caspase-9 and consecutive apoptosis. Seems to
CC       bind ATP.
CC   -!- SUBUNIT: Interacts strongly with caspase 2, weakly with caspase 9
CC       and with APAF1 in a cytochrome c-inducible way leading to the
CC       formation of an apoptosome. This interaction may be ATP-dependent.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=4;
CC       Name=1; Synonyms=NAC beta, DEFCAP-L;
CC         IsoId=Q9C000-1; Sequence=Displayed;
CC       Name=2; Synonyms=NAC alpha, DEFCAP-S;
CC         IsoId=Q9C000-2; Sequence=VSP_004327;
CC       Name=3; Synonyms=NAC gamma;
CC         IsoId=Q9C000-3; Sequence=VSP_004326, VSP_004327;
CC       Name=4; Synonyms=NAC delta;
CC         IsoId=Q9C000-4; Sequence=VSP_004326;
CC   -!- TISSUE SPECIFICITY: Widely expressed. Isoforms 1 and 2 are
CC       expressed in peripheral blood leukocytes, chronic myelogenous
CC       leukemia cell line K-562, followed by thymus, spleen and heart.
CC       Also detected in lung, placenta, small intestine, colon, kidney,
CC       liver and muscle.
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 1 CARD domain.
CC   -!- SIMILARITY: Contains 7 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF298548; AAG15254.1; -.
DR   EMBL; AF310105; AAG30288.2; -.
DR   EMBL; AF229059; AAK00748.1; -.
DR   EMBL; AF229060; AAK00749.1; -.
DR   EMBL; AF229061; AAK00750.1; -.
DR   EMBL; AF229062; AAK00751.1; -.
DR   EMBL; AB023143; BAA76770.1; -.
DR   PIR; T17255; T17255.
DR   HSSP; P13489; 1A4Y.
DR   MIM; 606636; -.
DR   GO; GO:005622; C:intracellular; IC.
DR   GO; GO:016606; F:apoptosis activator activity; NAS.
DR   GO; GO:0008656; F:caspase activator activity; NAS.
DR   GO; GO:0019899; F:enzyme binding activity; IPI.
DR   GO; GO:0006919; P:caspase activation; NAS.
DR   GO; GO:0006917; P:induction of apoptosis; NAS.
DR   InterPro; IPR001315; CARD.
DR   InterPro; IPR000767; Disease_resist.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF00560; LRR; 2.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   PRINTS; PR00364; DISEASERSIST.
DR   PROSITE; PS50209; CARD; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Apoptosis; ATP-binding; Leucine-rich repeat; Repeat;
KW   Alternative splicing.
FT   DOMAIN        1     92       DAPIN.
FT   DOMAIN      328    637       NACHT.
FT   REPEAT      704    725       LRR 1.
FT   REPEAT      807    830       LRR 2.
FT   REPEAT      864    887       LRR 3.
FT   REPEAT      921    944       LRR 4.
FT   REPEAT      950    973       LRR 5.
FT   REPEAT     1199   1215       LRR 6.
FT   REPEAT     1216   1236       LRR 7.
FT   DOMAIN     1374   1463       CARD.
FT   NP_BIND     334    341       ATP (POTENTIAL).
FT   VARSPLIC    958    987       Missing (in isoform 3 and isoform 4).
FT                                /FTId=VSP_004326.
FT   VARSPLIC   1262   1305       Missing (in isoform 2 and isoform 3).
FT                                /FTId=VSP_004327.
FT   MUTAGEN     340    340       K->L: NO EFFECT.
FT   MUTAGEN     340    340       K->S: NO EFFECT.
```

## Left column

```
FT   CONFLICT     155    155    L -> H (IN REF. 1).
FT   CONFLICT     246    246    T -> S (IN REF. 1).
FT   CONFLICT     782    782    T -> S (IN REF. 1).
FT   CONFLICT     878    878    T -> M (IN REF. 1).
FT   CONFLICT     995    995    N -> V (IN REF. 1).
FT   CONFLICT    1119   1119    N -> I (IN REF. 1 AND 6).
FT   CONFLICT    1184   1184    N -> L (IN REF. 1).
FT   CONFLICT    1241   1241    V -> L (IN REF. 1).
FT   CONFLICT    1366   1366    F -> C (IN REF. 1).
SQ   SEQUENCE   1473 AA;  165865 MW;  438F0DCE45C2562D CRC64;
```

Query Match 23.7%; Score 762; DB 1; Length 1473;
Best Local Similarity 33.2%; Pred. No. 1e-47;
Matches 226; Conservative 103; Mismatches 207; Indels 132; Gaps 22;

```
Qy     1  MLAQPQRLFILDGADEL-P--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV   58
          ::::|:||||||||| |      |: ||  ||||||||||  :||||||| |:|  :|:
Db   398  ILSRPERLLFILDGVDE-PGKVLQEPSSELCLHWSQPQPADALLGSLLGKTLPEASFLI   456

Qy    59  TTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFKFFRDERRAERAYRFVKENETLFAL   118
          | | :|  ::|  | | :  || |||| : ::||| |:||| ||:|| ||:|:::| ||
Db   457  TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSNKELWAL   516

Qy   119  CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLLSSAPVADGPRLQGDLRN  178
          | ||:| |:.||.| ||::  |:||:||||| : |:| :|  :|: |:  :  :::-LRD
Db   517  CLVPWVSWLACTCLMQQMRKREKLTLTSKTTTTLCLHYLAQALQAQPL--GPQ----LRD   570

Qy   179  LCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETE--VTYQFIDQS   235
          || ||:||: |:: ::  ::||::| | :  :|||:     | | :      : | |:|
Db   571  LCSLAAEGIWQKKTLFSPDDLRKHGLDGAIIST-FLKM---GILQEHPIPLSYSFIHLC   625

Qy   236  FQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVL----------------T   277
          ||||:||:||:||               ::|          | |||
Db   626  FQEFFAAVSYVLED--------EKGRG-------KHSNCIIDEKTLEAYGHGLFGAST   670

Qy   278  TRFLFGLLSAERXRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLED   337
          ||||:|||| :     | :  |  | : : :       :|||
Db   671  TRFLLGLLSDEGEREMENIFHCRLSQ-GRNLMQWV------------FSLQLL       720

Qy   338  EEPEEEEGEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVAFC--RXDVAVLS    394
          :  :  ||  :  | :: ||||: : :|     |:    :::  :    ||:::||:
Db   711  LQP---------HSLESLHCLYETRNKTFLTQVMAHFEEMGX------CVETDMELLVCT  755

Qy   395  YCVRCCPAGQALRLISCR-----------LVAAQEKKKKSLGKRLQ               429
          :|::         : | |             |:| :||:||||:|
Db   756  FCIKFSRHVKKLQLIEGRCHRSTWSPTMVVLFRWVPVTDAYKGILFSSVLKVTRNL-KELD  814

Qy   430  ASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD-SEA-RAAP  489
           | ||:     | | : |  |:  :  |   || ||| |:  |  : |:: |:  ||:|
Db   815  LS---GNS-----LSHSAVKSICKTLRRPRCLLETLRLAGGGLTAEDCKELAFG-RANQ   865

Qy   490  ALTELGLHNRLSEAGLRMLSEGLAWPQCRVQCTVRVQLPDPQRGL----QYLVGMLRQS  544
           |||| | :.||:||:::|  |  : :|::| :    :    :|      | | :|:|:
Db   866  TLTELDLSFNVLTDAGAKHLCQRLRQPSCKLQ--RLQL--VSGLSTSDCCQLASVLSAS   921

Qy   545  PALTTLDLSGCQLPAPMVTYLZAVLQHQGCGLQTLSLASVELSEQSLQELQAVKRAKPDL  604
          |:|  |||   ::  :::  |  :|:|  | |  | |  : |:: |::||:| |:|||:|
Db   922  PSLKELDLQQNNLDDVGVRLLJEGLRHPACKLIRLGLDQTTLSDEMRQELRALEQEKPQL  981

Qy   605  VI  606
          :|
Db   982  LI  983
```

RESULT 8
PYA3_HUMAN
```
ID   PYA3_HUMAN     STANDARD;      PRT;   980 AA.
AC   Q8WX94;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   PYRIN-containing APAF1-like protein 3.
GN   PYPAF3.
```

## Right column

```
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=22162427; PubMed=12219269;
RA   Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA   Lora J.M., Geddes B.J., Briskin M., DiStefano P.S., Bertin J.;
RT   "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT   activation of NF-kappa B and caspase-1-dependent cytokine
RT   processing.";
RL   J. Biol. Chem. 277:29874-29880(2002).
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 9 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF464765; AAL69963.1; -.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat.
FT   DOMAIN       172     93       DAPIN.
FT   DOMAIN             491       NACHT.
FT   REPEAT       614    638       LRR 1.
FT   REPEAT       674    697       LRR 2.
FT   REPEAT       760    784       LRR 3.
FT   REPEAT       788    810       LRR 4.
FT   REPEAT       817    840       LRR 5.
FT   REPEAT       845    868       LRR 6.
FT   REPEAT       874    897       LRR 7.
FT   REPEAT       902    928       LRR 8.
FT   REPEAT       933    957       LRR 9.
FT   NP_BIND      178    185       ATP (POTENTIAL).
SQ   SEQUENCE   980 AA;  111806 MW;  822AF2FD4338003D CRC64;
```

Query Match 20.2%; Score 650; DB 1; Length 980;
Best Local Similarity 28.6%; Pred. No. 9.6e-40;
Matches .221; Conservative 102; Mismatches 256; Indels 194; Gaps 23;

```
Qy     1  MLAQPQRLFILDGADELPALGGPEAAPCTDPFEAASGARV-LGGLLSKALLPTALLLVT   59
          :|:||:|:|:||| || | ::|:  |  :|:           |||||| :|::| |||:
Db   241  ILAAQRILFVVDGLDELKPPGALIQDICGDWEKKKPVPVLLGSLLKRKMLPRAALLVT   300

Qy    60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFKFFRDERRAERAYRFVKENETLFALC  119
          ||  :  ||: ||  |:  :|  |  :|:    : : :||||| |::   :::  | | 
Db   301  TRPRALRDLQLLAQQPIYVRVEGFLEEDRRAYFLRHFGDEDQAMRAFELMRSNAALFQLG  360

Qy   120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLLSSAPVADGPRLQGDLRNL  179
             :|||||||.|::|:| | :   :|   |  :|||: ||:|:|:|||     ::| ::|
Db   361  SAPAVCWIVCTTLKLQMEKGEDPVPTCLRTGLFLRFLCS-----RFPQGAQLRGALRTL  415

Qy   180  CRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF  239
          | |:  |:  :::  : : |:: |:  |::  :|:        | ::    | |:|||:|
Db   416  SLLAAQGLWAQMSVFHREDLERLGVQESDLR-LFLDGDILRQDRVSKGCYSFIHLSFQQF  474

Qy   240  LAALSYLLE-DGGVPRTA---AGGVGTLLRGDAQ-PHSHLVLTTRFLFGLLSAERMRDI  293
          |:||:| ||   |           :| :|:|     ::  :      :||| :|: :::
Db   475  LTALFYALEKEEGEDRDGHAWDIGDVQKLLSGEERLKNPDLIQVGHFLFGLANEKRAKEL  534

Qy   294  ERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEPNYPL   353
```

```
Db  535  EATFGCRMSPDIKQELLQ------------C---------------------------KAHHANKP_  563
                 |:||  ::|||  || |
Qy  354  -----ELLYCLYETQEDAFVRQALCRFPELAL-----QRVRFCR-----MDVAVLSYC  396
                 :|:  :  ||| : ::|
Db  564  SVTDLKEVLGCLYESQEELAKVVVAPFKEIS:HLTNTSEVHCSPSLKHCQDLQKLSLQ  623
                 :|  | |  :|   :|  |
Qy  397  V---------------RC----CP--AGQALRLIS-----CRLVAQEKKK-----  421
                 |
Db  624  VAKGVFLENYMPFELDEFEKCTYLTIPNWARQDLRSLRLWTDFCSLFSSNSNLKFLEVK  683
                 |||  ||   :||   :|  |  ||   ||  |||
Qy  422  -------------------------KSLGKR--LQASLGG---  434
                             |||  :||
Db  684  QSFLSDSSVRILCCHVTRSTCHLQKVEIKNVTPDTAYRDFCLAFIGKKTLTHLTLAGHIE  743
                 |:  |:  :  : ||  || :|
Qy  435  ------------------------GSSQGTKQ-------------LPASLLHP  451
                                          |::|
Db  744  WERTMYLMLCDLLRNHKCNLQYLRLGGHCATPEQWAEFFYVLKANQSLKHLRLSANVLLD  803
                 |||  :|   ||   |   :|  |  :|  |
Qy  452  ------LFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPA:TELGLHNRLSEAGL  506
                 |:||  |    :||  :|  | |:
Db  804  EGAMLLYKTMTRPKHFLQMLSDENCRLTEASCKDLAAVLVVSKKLTHLCLAKNFPIGDTGV  863
                 ||  |  ||  :|::||  : |  |  :|
Qy  507  RYLSEGLAWPQCRVQT-VRVQLPDPQRGLQY:VGMLRQSPALTTLDLSGCQLPAPMVTY_  565
                 | |  |::  | |   :|:|  | |  ||
Db  864  KFLCEGLSYPDCKLQTLVLQQCSITKLGCRYLSEALQEACSITNLDLSINQI-ARGIWIL  922
                 ||  |||  || :|:|  |  |::|  :|  |:  |::|
Qy  566  CAVLQHQGCGLQTLSLASVEISEQSLQELQAVKRAKPDLVITHPALDGHPQPP  619
                 |   | |  || :||     |:|  ::| :| |
Db  923  CQALENPNCNLKHLRLKTYEINLEIKKLLEEVKEKNPKLTIDCNA-SGATAPP  974
```

RESULT 9

```
NAL2_HUMAN
ID   NAL2_HUMAN     STANDARD;      PRT;  1062 AA.
AC   Q9NX02; Q9VN5; Q9H6G6; Q9HAV9; Q9NWK3;
DT   16-OCT-2001 (Rel. 40, Created)
DT   16-OCT-2001 (Rel. 40, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   NACHT-, LRR- and PYD-containing protein 2 (Nucleotide-binding site
DE   protein 1) (PYRIN-containing APAF1-like protein 2).
GN   NALP2 OR NBS1 OR PYPAF2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 2).
RX   MEDLINE=21169419; PubMed=11270363;
RA   Bertin J., DiStefano P.S.;
RT   "The PYRIN domain: a novel motif found in apoptosis and inflammation
RT   proteins.";
RL   Cell Death Differ. 7:1273-1274(2000).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=21148093; PubMed=11250163;
RA   Martinon F., Hofmann K., Tschopp J.;
RT   "The pyrin domain: a possible member of the death domain-fold family
RT   implicated in apoptosis and inflammation.";
RL   Curr. Biol. 11:R118-R120(2001).
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RX   MEDLINE=22162427; PubMed=12019269;
RA   Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA   Lora J.M., Geddes B.J., Briskin M., DiStefano P.S., Bertin J.;
RT   "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT   activation of NF-kappa B and caspase-1-dependent cytokine
RT   processing.";
RL   J. Biol. Chem. 277:29874-29880(2002).
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Colon, and Kidney epithelium;
RA   Watanabe K., Kumagai A., Itakura S., Yamazaki M., Tashiro H., Ota T.,
RA   Suzuki Y., Obayashi M., Nishi T., Shibahara T., Tanaka T.,
```

```
RA   Nakamura Y., Isogai T., Sugano S.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2000) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Lung, and Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: May be implicated in apoptosis (By similarity).
CC   -!- COFACTOR: Binds ATP (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q9NX02-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q9NX02-2; Sequence=VSP_005522;
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 9 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF298547; AAG15253.1; ALT_INIT.
DR   EMBL; AF310106; AAG30289.1; -.
DR   EMBL; AF464764; AAL69962.1; -.
DR   EMBL; AK000517; BAA91223.1; -.
DR   EMBL; AK025952; BAB15293.1; -.
DR   EMBL; AK000784; BAA91377.1; ALT_INIT.
DR   EMBL; BC003592; AAH03592.1; -.
DR   EMBL; BC001039; AAH01039.1; -.
DR   HSSP; P10775; 2BNH.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Apoptosis; ATP-binding; Leucine-rich repeat; Repeat;
KW   Alternative splicing.
FT   DOMAIN        1     94       DAPIN.
FT   DOMAIN      207    526       NACHT.
FT   REPEAT      467    491       LRR 1.
FT   REPEAT      622    645       LRR 2.
FT   REPEAT      754    777       LRR 3.
FT   REPEAT      810    832       LRR 4.
FT   REPEAT      839    862       LRR 5.
FT   REPEAT      867    890       LRR 6.
FT   REPEAT      924    947       LRR 7.
```

```
FT   REPEAT        981   1005       LRR 8.
FT   REPEAT       1010   1033       LRR 9.
FT   NP_BIND       218    220       ATP (POTENTIAL).
FT   DOMAIN        518    523       POLY-GLU.
FT   VARSPLIC      133    154       Missing (in isoform 2).
FT                                  /FTId=VSP_005522.
FT   CONFLICT        1      1       M -> V (IN REF. ).
FT   CONFLICT       35     35       L -> P (IN REF. 2).
FT   CONFLICT      304    304       J -> S (IN REF. 4; BAB15293).
FT   CONFLICT      364    364       E -> K (IN REF. 1).
FT   CONFLICT      980    980       MISSING (IN REF. 1).
FT   CONFLICT     1052   1052       A -> E (IN REF. 5).
SQ   SEQUENCE     1062 AA;   120514 MW;  4DBB0F6E9C2BC8A7 CRC64;

Query Match             19.2%;   Score 616.5;  DB 1;  Length 1062;
Best Local Similarity   27.3%;   Pred. No. 3e-37;
Matches  204;   Conservative  104;   Mismatches  260;   Indels  179;   Gaps  :8;

Qy      1 MLAQPQRLFLFILDGADELPA..GGPEAAPCTDPPEAASGARV.LGGLLSKALLP.ALL.VT   59
          :|||:| :||||:|:::||: :|:|:::||:|    ||| || :| ||||||
Db    276 ILAQARKILFVIDGFDELGAAPGALIEDICGDWEKKKPVPVLLGSLLNRVMLPKAALLVT  335

Qy     60 TRAAAPGRLQGRLCSPOCAEVRGFSDKDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC  119
          |   ||| ||| | :  |  ||    |   |  ||  |:|  ||:||:  |: |: ||:
Db    336 TRPRALRDLRILAEEPIYIRVEGFLEEDRRAYFLRHFGDEDQAMRAFELMRSNAAL.FCLG  395

Qy    120 FVPFVCVK.VCTVLRQQLELGRDLSRTSKTTTSVYLLFTSVLSSAPVADGPRLQGDLRNL.179
          | |:||  :|||||   | ||:  :| : |: | :: |  ||: :||:||||||:||| :
Db    396 SAPAVCW.VCTLLKLQVEKGEDFVPTCLTRTGLFLRFLCS----.RFPQGAQLRGALRTL  450

Qy    180 CRLAREGVLGRRAQFAEKELEQELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQEF    239
          |:|| :|:| |:  ||| ||:| |::|    |:|  | || |:|:::|  ||: |
Db    451 SLLAAQGLWAQTSVLHREDLER:GVQESDLR.LFLDGDILRQDRVSKGCYSFIH:LSFCQF  509

Qy    240 LAALSYLLE-------DGGVFRTAAGGVGTLRG-DAQPHSHLVLTTRFLFGLLSAERMR   291
          | | ||:| |      :|:| || |  |:|||   |:|  | |  | ||| :| :|:|
Db    510 LTALFYTLEKEEEDRDGHT-WDIGDVQKLLSGVERLRNPDLIQAGYYSFGLANEKRAK    567

Qy    292 DIERHFGCWVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEEGEEPNY   351
          :||:||   |  || ||:|| |    |      | |  |:|||:|: | :
Db    568 ELEATFGCRMSPDIKQELLRC----DISCKGGHSTVTD----LQ---------------  603

Qy    352 PLELLYCLYETOEDAFVRQALCRFPELALQRVFRCRMDVAVLSYCVRCCPAGQAIRL---   408
          |  ||| :|| |    |  |  |||:| | |::  |  |  | |  |||||||:|:|
Db    604 --ELLGCLYESQEELVKEVWAQFKE:SL---HLNAVDVVPSSFCVKHCRNLQKVSLQVI  658

Qy    409 -----------------------------------------------I            409
                                                         |
Db    659 KENL.PENVTASESDAEVERSCDDQHMLPFWTDLCSIFGSNKDLMGLAINDSFLSAS:VRI  718

Qy    410 SCRLVRA------CEKKKSLG----KRLQASLGGGSS------CGTTK--QLPA---SL   448
          :|:    |      :|::|        |:| |  |             |  :||:   ||
Db    719 LCEQIASD:CHLQRVVFKNISPADAHRNLCLALRGHK:VTY:CLQGNDQCDMFPALCEVL  778

Qy    449 LHP------------------------------------------LFQAVTDP        459
          |                                           |  :| ||
Db    779 RHPECNLRYLGLVSCSATTQQWADLSLALEVNQSLTCVNLSNELLDEGAKLLYT:LRHP  838

Qy    460 LCH:SSE:TLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCR  519
          |  :|::| ::|||     |:|:||||:  ||||:|||||||||:|||:|:||:|| ||:
Db    839 KCFLQRLSLENCHLTEANCKDLAAVLVVSRELTH:CLAKNPIGNTGVKFLCEGLRYFECK  898

Qy    520 VQTVRVQLPD-PQRGLQYLVGM.RQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQ:  578
          :|:          ||||||||| :|:|||||||||| ||  |:|||||||||| ||:|
Db    899 LQTIVLWNCDITSDGCCDLTKLLQEKSSLLCLDLGLNHIGVKGMKFLCEALRKPLCNLRC  958

Qy    579 :SLASVELSEQSLQELQAVKRAKPDLV  605
          :| |:  :|||::| ||| |::||||
Db    959 LWLWGCSIPPFSCEDLCSALSCNQSLV  985
```

RESULT 10
NAL4_HUMAN

ID   NAL4_HUMAN     STANDARD;     PRT;    994 AA.
AC   Q96MK2; Q96AY6;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   NACHT-, LRR- and PYD-containing protein 4 (PAAD and NACHT-containing
DE   protein 2) (PYRIN-containing APAF1-like protein 4) (Ribonuclease
DE   inhibitor 2).
GN   NALP4 OR PYPAF4 OR PAN2 OR RNH2.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Martinon F., Tschopp J.;
RT   "NALP4 a novel member of the PYD, NACHT, and LRR family.";
RL   Submitted (NOV-2001) to the EMBL/GenBank/DDBJ databases.
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Fiorentino L., Reed J.C.;
RT   "Pan2, a novel PAAD-containing protein.";
RL   Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
RN   [3]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Bertin J.;
RT   "PYPAF4: a novel PYRIN-containing APAF1-like protein.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [4]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RA   Miyamoto T.;
RT   "NEDO human cDNA sequencing project.";
RL   Submitted (FEB-2002) to the EMBL/GenBank/DDBJ databases.
RN   [5]
RP   SEQUENCE FROM N.A. (ISOFORM 3).
RA   Ninomiya K., Wagatsuma M., Kanda K., Kondo H., Yokoi T., Kodaira H.,
RA   Furuya T., Takahashi M., Kikkawa E., Omura Y., Abe K., Kamihara K.,
RA   Katsuta N., Sato N., Tanikawa M., Yamazaki M., Sugiyama T., Irie R.,
RA   Otsuki T., Sato H., Wakamatsu A., Ishii S., Yamamoto J., Isono Y.,
RA   Kawai-Hio Y., Saito K., Nishikawa T., Kimura K., Yamashita H.,
RA   Matsuo Y., Nakamura Y., Sekine M., Kikuchi M., Murakawa K.,
RA   Karehori K., Takahashi-Fujii A., Oshima A., Sugiyama A., Kawakami B.,
RA   Suzuki Y., Sugano S., Nagahari K., Masuho Y., Nagai K., Isogai T.;
RL   Submitted (OCT-2001) to the EMBL/GenBank/DDBJ databases.
RN   [6]
RP   SEQUENCE OF 436-994 FROM N.A. (ISOFORM 2).
RC   TISSUE=Placenta;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Warg J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.C.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Holyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=3;
CC         Comment=Experimental confirmation may be lacking for some
CC         isoforms;
CC       Name=1;

```
Qy   178 NLCRLAREGVLGRRAQPAEKELEQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQSFQ 237
            |||| :|: :  || ::| :  |  : ||:  |:||| ::: |||:|||:|||:|: :
Db   398 ALCSLAAEGMWTDTFEFCEDDLRRNGVVDADIPAL-LGTKILLKYGERESSYVFLHVCIQ 456

Qy   238 EFLAALSYLLEDG-GVPRTAAGGVGTLLRGDAQP--HSHLVLTTRFLFGLLSAERMRDIE 294
            || || :||    |   :  |:|| ||  |:     :    :|| | |||:  :| :
Db   457 EFCAALFYLLKSHLDHPHPAVRCVQELLVANFEKARRAHWIFLGCFITGLLNKKEQEKLD 516

Qy   295 RHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEGEGEPNYPLS 354
            ::|| :|: :||:
Db   517 AFFGFQLSQEIKQ-----------------QIHQCLKSLGERGNPQGQVDS------LA 552

Qy   355 LLYCLYETQEDAFVRQALCRFPBLALQRVRF---CRMDVAVLSYCVRCCPAGQALRLISC 411
            :::||:| ||:||| ||  :  |  |:  |      : | | :| :     :|||||:|
Db   553 IFYCLFEMQDPAFVKQAV-----NLLQEANFHIDNVDLVVSAYCLKYCSSLRKL-CFSV 606

Qy   412 RLVAAQEKKKKSLGK------ 426
            ::: : |
Db   607 QNVFKKEDEHSSTSDYSLICWHICSVLTTSGHLRELQVQDSTLSESTFVTWCNQLRHPS 666

Qy   427 -RLQ-----ASLGGGS-------SQGTTKQLPASL------LHPLFQAMTDPLCHLS 464
            :||       |: ||        :|| | ||| |       ||||| | |||| ||
Db   667 CRLQKLGINNVSFSGQSVLLFEVLFYQPDLKYLSFFLTKLSRDDIRSLCDALNYPAGNVK 726

Qy   465 SLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRISEAGLRMISEGLAWPQCRVQTVR 524
            .| : || | | | ::| || |||||||:|||| || |||||::| | || : | :|
Db   727 ELALVNCHLSPIDCEVIAGLLTNNKKLTYLNVSCNQLITGVPLICEAL---CSPDTVL 781

Qy   525 VQL----PDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQCGLQTL 579
            | :     ||:| | ||| || ||||:|||||:||||||||:||||| :|| | ||:|
Db   782 VYLMLAFCHLSEQCCEYISEMLLRNKSVRYLDLSANVLKDEGLKTLCEALKHPDCCLDSL 841

Qy   580 --------------SLASVE-SEQSLQELQ 595
                          :|| |  ||| :||:|
Db   842 CLVKCFITAAGCEDLASALISNQNLKILQ 870
```

```
RESULT 11
MATER_HUMAN
ID   MATER_HUMAN    STANDARD;      PRT;   1200 AA.
AC   P59047;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Mater protein homolog.
GN   MATER.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   TISSUE=Ovary;
RX   MEDLINE=21922687; PubMed=11925379;
RA   Tong Z.-B., Bondy C.A., Zhou J., Nelson L.M.;
RT   "A human homologue of mouse Mater, a maternal effect gene essential
RT   for early embryonic development.";
RL   Hum. Reprod. 17:903-911(2002).
CC   -!- TISSUE SPECIFICITY: Oocyte-specific.
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
CC   -!- CAUTION: It is not obvious that this is the ortholog of mouse
CC       Mater.
```

```
CC        IsoId=Q96MN2-1; Sequence=Displayed;
CC        Name=2;
CC        IsoId=Q96MN2-2; Sequence=VSP_003917;
CC        Name=3;
CC        IsoId=Q96MN2-3; Sequence=VSP_003916;
CC   -!- SIMILARITY: Contains 1 DAPIN domain.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 8 leucine-rich (LRR) repeats.
CC   -!- CAUTION: A stop codon in Ref.4 was read through in position 41 to
CC       extend the sequence.
```

```
DR   EMBL; AF442488; AAL35293.1; -.
DR   EMBL; AY072792; AAL68396.1; -.
DR   EMBL; AF479747; AAL87104.1; -.
DR   EMBL; AF482706; AAL88672.1; ALT_INIT.
DR   EMBL; AK056688; BAB71254.1; -.
DR   EMBL; BC016443; AAH16443.1; -.
DR   InterPro; IPR007091; LRR_RNrich.
DR   InterPro; IPR007211; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   Repeat; Leucine-rich repeat; Alternative splicing.
FT   DOMAIN       1     94       DAPIN.
FT   DOMAIN     149    472       NACHT.
FT   REPEAT     637    660       LRR 1.
FT   REPEAT     698    721       LRR 2.
FT   REPEAT     722    745       LRR 3.
FT   REPEAT     750    777       LRR 4.
FT   REPEAT     806    833       LRR 5.
FT   REPEAT     863    886       LRR 6.
FT   REPEAT     920    943       LRR 7.
FT   REPEAT     949    972       LRR 8.
FT   VARSPLIC     1     93       MAASFSDFGLWYLEELKEEFRKPKEHLKQMTLQLELKQ
FT                               IPWTEVKKASREELANLLIKGYEEQQAWNITLRIFQKMDRK
FT                               DLCMKVMRERT -> MQECLTLWVFSPLALTDS (in
FT                               isoform 3).
FT                               /FTId=VSP_003916.
FT   VARSPLIC   731    786       Missing (in isoform 2).
FT                               /FTId=VSP_003917.
FT   CONFLICT    55     55       L -> P (IN REF. 4).
FT   CONFLICT    99     99       Y -> H (IN REF. 4).
FT   CONFLICT   177    177       I -> T (IN REF. 4).
FT   CONFLICT   209    209       W -> R (IN REF. 4).
FT   CONFLICT   296    296       I -> V (IN REF. 4).
FT   CONFLICT   720    720       Y -> C (IN REF. 4).
FT   CONFLICT   957    957       L -> P (IN REF. 4).
SQ   SEQUENCE   994 AA;  113444 MW;  1C8CFC68B18398C9 CRC64;
```

```
Query Match:      18.6%;       Score 598.5;  DB 1;   Length 994;
Best Local Similarity  28.0%;  Pred. No. 5.6e-36;
Matches  193;  Conservative 114;  Mismatches 251;  Indels 131;  Gaps  19;

Qy     1 MLAQPQRLFLFILDGADELP-ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT 59
            :  ||||| :|| | ||: ||||| |||  |  ||    ||::|| ||||| |||: :
Db   219 IVSQPERLLFVIDSFEELQGGLNEPDSDLCGDLMEKRPVQVLLSSLLRKKMLPEASLLIA 278

Qy    60 TRAAAPGRLQGRLCSPCCAEVRGFSDKDKKYFYKFPDERRAERAYRFVKENETLFALC 119
            :  | ||||| || : |  ||:|| | :| ||||   |||   | : || :||||:|
Db   279 IKPVCPKELRDQVTISEIYQPRGFNESDRLVFCCFFKDPKRAMEAFNLVRESEQLFSIC 338

Qy   120 FVPFVCKIVCTVLRQLELGRDLSRTSKTTTSVYLFITSVLSSAPVADG--PRLQGDLR 177
            :|| :| :|||:|:|:: ||||: || || ||||:  |:|||||| |    ||:|| |
Db   339 QIPLLCWLCTSLKQEMQKGKDLALTCQSTTSVYSSSVFN-LFTPEGAEGTTPQTQHQLK 397
```

```
CC   -----------------------------------
DR   EMBL; AY054986; AAL15549.1; -.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNint.
DR   InterPro; IPR003590; LRR_RNint_sub.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   InterPro; IPR004020; PAAD_DAPIN_dom.
DR   Pfam; PF00560; LRR; 3.
DR   Pfam; PF02758; PAAD_DAPIN; 1.
DR   PRINTS; PR00019; LEURICHRPT.
DR   SMART; SM00368; LRR_RI; 11.
DR   PROSITE; PS50824; DAPIN; 1.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat.
FT   DOMAIN       57    148       DAPIN.
FT   DOMAIN      280    602       NACHT.
FT   REPEAT      704    727       LRR 1.
FT   REPEAT      730    753       LRR 2.
FT   REPEAT      780    803       LRR 3.
FT   REPEAT      809    832       LRR 4.
FT   REPEAT      836    863       LRR 5.
FT   REPEAT      865    892       LRR 6.
FT   REPEAT      893    916       LRR 7.
FT   REPEAT      950    973       LRR 8.
FT   REPEAT      979   1002       LRR 9.
FT   REPEAT     1007   1034       LRR 10.
FT   REPEAT     1034   1059       LRR 11.
FT   REPEAT     1064   1092       LRR 12.
FT   REPEAT     1121   1142       LRR 13.
FT   NP_BIND     286    293       ATP (POTENTIAL).
SQ   SEQUENCE   1200 AA;  134235 MW;  9A070D2A771B28FA CRC64;
```

```
Query Match             15.8%;  Score 507.5;  DB 1;  Length 1200;
Best Local Similarity   26.7%;  Pred. No. 3.1e-29;
Matches  187;  Conservative  98;  Mismatches 299;  Indels 117;  Gaps  14;

Qy        1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPPEAASGARVLGGLLSKALLPTALLLVTT    60
              :  ::|||||:|  | :  :| :: |   |:: |:: |:|| | || :|:  :| ||::
Db      350 :XSRPERLLFIIIDGFDD:GSVLNDTKLCKDWAEKQPPFTI:IRSLLRKVLLPESFLIVTV   409

Qy       61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFPRDCERRAERAYRFVKENET:FALCF   120
              | |:||| : :| :| | :| ||||:|   :||::   |::| |  |: |: |:| ||
Db      410 REVGTEKLKSEVVSPRYLLVFGIISGEQRIHSLLERG:GEHQKTCG:LRAIINNREL:DQCQ   469

Qy      121 VPFVCWIVCTVLRQQLELGRCLSRTSKTTTSVYLLFTTSVLSSAPVAD---GPR:QCGLR   177
             |::   |  ||   |  |    ::  ::  :|:|  :: | :: : ||     ||  :|:|
Db      470 VPAVGSLICVALQLQDEVVGES:VAPFNQTLTGLJAAFAFHQL:PRGVVRRCLNLEERVVZK   529

Qy      178 NLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSKKELPGV:ETEVTYCFIDCSFQ    237
             | :|  |:  |  |   ||| :|    |::| | |   ||  |:| : |: | :|  :|::
Db      530 RFCRMAVEGVWNRKSVFDGDIL:VVCGLGSELRALFHWNILLPD-SHCEEYTFFHLS:Q    598

Qy      238 EFLAALSYLLEDGGV-----PRTAAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRD   292
             :  || :|: ||     ::: | |  :| |  | | |:: | | |  ||  | :|::::|
Db      589 DFCAALYYVLEGLEIEPALCFLYVEKTKRSMELKQAGFHIHSLWMKRF:FGLVSEDVRRP   648

Qy      293 IERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEEEGEEPNYP   352
             :  ||| |   :|   | :| |     || |    |  | |:  ::| :|:: :| ||
Db      649 LEVLLGCFVPLGVKQKLLHWVSLLGQ----QPNAT------------------TPGDT   684

Qy      353 LELLYCLYETQEDDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAGQALRL----   408
              : : :|| :|:|| |: |   | | |  ||    :: |::| :| |  :|| |:
Db      685 :DAFHCLFETQDKPFVRLALNSFQEVWLPINQ--NLDLIASSFCL:QHCPYLRKIRVDVKG   742

Qy      409 ------ISCRLVAAQEKKKKSLGKRLQ---------ASLGGGSS---QGTTKQLP      445
                    : :||:| :|  :: ||  :             : |:     |||: ||
Db      743 IFPRDESAEACPVVPLMWRDKTLIEEQWEDFCSMLGTHPHLRQLDLGSSILTERAMKTLC   802

Qy      446 ASLLHP-----:LF-----------------------------------QAMT        457
                 ::|:       ::
Db      803 AKLRHPTCKIQTLMFRNAQITPGVVQHLMRIVMANRNLRSLNLGG:THLKEEDVRMACEALK   862

Qy      458 DPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTE:GLLHNRLSEAGLRM:SEGLAWPQ   517
```

```
Db      863 HPKCLLESLR:DCCGLTHACYLKISQILTTSPSLKSLSLAGNKVTDQGVTPLSDALRVSQ   922
                ||  |::| :|  |  | ::| : :: |  ||| :|::| |::| ||| : |: | :||
Qy      518 CRVQTVRVQLPD--PQRGLQYLVGMLRQSPALTT:DLSGCQLPAPMVTYLCAVLQHQGC   574

Db      923 CALQ--KLILEDCGITATGCQSLASALVSNRSLTHLCLSNNSLGNEGVNLLCRSXRLPHC   980
                 ||:   ||  :|  | |  :|:  || :||:||  |  || || || |||  |  ||
Qy      575 GLQTLSLASVELSEQSLQELQAVKRAKPDLVITHPALDGHP   615

Db      981 SLQRLMLNQCHLDTAGCGSLALAMGNSWL--TH:SLSMNP   1019
```

```
RESULT 12
MATE_MOUSE
ID   MATE_MOUSE     STANDARD;     PRT;  1111 AA.
AC   Q9RIMS; Q9JLR2;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   28-FEB-2003 (Rel. 41, Last annotation update)
DE   Maternal antigen that embryos require (Mater protein) (Ooplasm-
DE   specific protein 1) (OP1).
GN   MATER.
OS   Mus musculus (Mouse).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
OX   NCBI_TaxID=10090;
RN   [1]
RP   SEQUENCE FROM N.A.
RC   STRAIN=NIH Swiss; TISSUE=Ovary;
RX   MEDLINE=99360614; PubMed=10433232;
RA   Tong Z.-B., Nelson L.M.;
RT   "A mouse gene encoding an oocyte antigen associated with autoimmune
RT   premature ovarian failure.";
RL   Endocrinology 140:3720-3726(1999).
RN   [2]
RP   SEQUENCE FROM N.A. AND VARIANTS ASP-7; LEU-42; LYS-87; PRO-256;
RP   GLN-977 AND 1003-ASN-ASN-1004.
RC   STRAIN=129/Sv;
RX   MEDLINE=20222768; PubMed=10754103;
RA   Tong Z.-B., Nelson L.M., Dean J.;
RT   "Mater encodes a maternal protein in mice with a leucine-rich repeat
RT   domain homologous to porcine ribonuclease inhibitor.";
RL   Mamm. Genome 11:281-287(2000).
RN   [3]
RP   FUNCTION.
RX   MEDLINE=20517328; PubMed=1062459;
RA   Tong Z.-B., Gold L., Pfeifer K.E., Dorward H., Lee E., Bondy C.A.,
RA   Dean J., Nelson L.M.;
RT   "Mater, a maternal effect gene required for early embryonic
RT   development in mice.";
RL   Nat. Genet. 26:267-268(2000).
CC   -!- FUNCTION: Necessary for embryonic development beyond the 2-cell
CC       stage.
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic.
CC   -!- TISSUE SPECIFICITY: Oocyte-specific.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 13 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; AF074018; AAD51762.1; -.
DR   EMBL; AF143573; AAF64393.1; -.
DR   EMBL; AF143559; AAF64393.1; JOINED.
DR   EMBL; AF143560; AAF64393.1; JOINED.
DR   EMBL; AF143561; AAF64393.1; JOINED.
DR   EMBL; AF143562; AAF64393.1; JOINED.
DR   EMBL; AF143563; AAF64393.1; JOINED.
```

```
DR   EMBL; AF143564; AAF64393.1; JOINED.
DR   EMBL; AF143565; AAF64393.1; JOINED.
DR   EMBL; AF143566; AAF64393.1; JOINED.
DR   EMBL; AF143567; AAF64393.1; JOINED.
DR   EMBL; AF143568; AAF64393.1; JOINED.
DR   EMBL; AF143569; AAF64393.1; JOINED.
DR   EMBL; AF143570; AAF64393.1; JOINED.
DR   EMBL; AF143571; AAF64393.1; JOINED.
DR   EMBL; AF143572; AAF64393.1; JOINED.
DR   PIR; A59000; A59000.
DR   HSSP; P1C775; 2BNH.
DR   MGD; MGI:1345193; Mater.
DR   GO; GO:0005829; C:cytosol; IDA.
DR   GO; GO:0009987; P:organogenesis; IMP.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   Pfam; PF00560; LRR; 3.
DR   PROSITE; PSS00837; NACHT; 1.
KW   ATP-binding; Leucine-rich repeat; Repeat; Polymorphism.
FT   DOMAIN        1    128   5 X APPROXIMATE TANDEM REPEATS.
FT   REPEAT        1     23   1 (INCOMPLETE).
FT   REPEAT       24     49   2.
FT   REPEAT       50     75   3.
FT   REPEAT       76    102   4.
FT   REPEAT      103    128   5.
FT   DOMAIN      191    513   NACHT.
FT   DOMAIN      640    664   LRR 1.
FT   REPEAT      691    714   LRR 2.
FT   REPEAT      737    762   LRR 3.
FT   REPEAT      776    799   LRR 4.
FT   REPEAT      804    832   LRR 5.
FT   REPEAT      834    856   LRR 6.
FT   REPEAT      861    884   LRR 8.
FT   REPEAT      890    913   LRR 9.
FT   REPEAT      918    941   LRR 10.
FT   REPEAT      950    974   LRR 11.
FT   REPEAT      975   1003   LRR 12.
FT   REPEAT     1005   1027   LRR 13.
FT   REPEAT     1032   1059
FT   NP_BIND     197    204   ATP (POTENTIAL).
FT   VARIANT       7      7   E -> D (IN STRAIN 129/SV).
FT   VARIANT      42     42   G -> L (IN STRAIN 129/SV).
FT   VARIANT      87     87   R -> K (IN STRAIN 129/SV).
FT   VARIANT     256    256   S -> P (IN STRAIN 129/SV).
FT   VARIANT     977    977   H -> Q (IN STRAIN 129/SV).
FT   VARIANT    1003   1004   SS -> NN (IN STRAIN 129/SV).
FT   SEQUENCE   1111 AA;  125502 MW;  4016A5D67A1C01F4 CRC64;

Query Match             14.6%;  Score 468.5;  DB 1;  Length 1111;
Best Local Similarity   23.8%;  Pred. No. 2e-26;
Matches  179;  Conservative 108;  Mismatches 270;  Indels 191;  Gaps  15;

Qy      1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVL-GGLLSKALLPTALLLVT   59
            ::|| |||||| :|| |   :     :||      ::|||  |  |||  | :
Db    260 IMSQPERLLFVIDGLDDMDSVLQHDDMTLSRDWKOEQPIYILMYSLLRKALLPQSFLIIT  319

Qy     60 TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC  119
Db    320 TRNTGLEKLKSMVVSPLYILVEGLSASRRSQLVLENISNESDRIQVFHSLIEHQLFDQC   379

Qy    120 FVPFVCWIVCTVLRQLELGRDLSRTSKTTTSVLLFITSVLSSAPVADGPRLCGD---L   176
Db    380 QAPSVSVCSLVCEALQLOKKLGKRCTLPCQTLTGLYATLVFHQLTLKRPSQSALSQEQITL  439

Qy    177 RNLCRLAREGVLGRRAQFAEKEIEQLELRGSKVQTLFLSKKELPGVLETEVTYQPIDQSF  236
Db    440 VGLCMCAAAEGVWTVRSVFYDDDLKONYSLKESEILALFHMNILLQVGHNSECCYVFSHLSL 499

Qy    237 QEFLAALSYLLEDGGVPRTAAGGV---------GTLRGDAQPHSH-VLTTRFLFG     283
Db    500 QDFFAALYYVLE--------GLEEWNQHFCFIENQRSIMEVKRTDDTRLLGMKRFLFG     549

Qy    284 LLSAERMRDIERHFGCMVSERVKQEALRWVQQGQGCPGVAPEVTEGAKGLEDTEPEEE     343
Db    550 LMNKDILKTLEVLFEYPVIPTVEQKLQHWVSLIAQQVNGTSPMDT---------          594

Qy    344 EEGEEPNYPLELLYCLYETOEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAG   403
Db    595 ---------LDAPYCLFESQDEEFVGGALKRFQEVWILLINQ-KMDLKVSSYCLKHCCNL   643

Qy    404 QALR-----LIS----CRLVAAQEKKK------SLGKRLQASLGGG-               435
Db    644 KAIRVDIRDLLSVDNTLELCPVVTVQETQCKPLLMEWWGNFCSVLGSLRNLKELDLGDSI   703

Qy    436 --------------SSQ------GTT-                                   441
Db    704 LSQRAMKILCLELRNQSCRIQKLTFKSAEVVSGLKHLWKLLFSNQNLKYLNLGNTPMKDD  763

Qy    442 --KQLFASLLHP---                                               451
Db    764 DMKLACEALKHPKCSVETLRLDSCELTIGYEMISTLLISTRLKCLSLAKNRVGVKSMI    823

Qy    452 -LFQAMTDPLCHLSSLT.SHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLS  510
Db    824 SLGNALSSSMCLLQKLILDNCGLTPASCHLLVSALFSNQNLTHLCLSNNSLGTEGVQQLC  883

Qy    511 EGLAWPQCRVQTV--RVQLPDPQRGLCYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCA   567
Db    884 QFLRNPECALQRLILNHCNIVDDAYG--FLAMRLANNTKLTHLSLTMNPVGDGAVKLLCE  941

Qy    568 VLQHOGGCGLQTLSLASVELSEQSLQEL                                 594
Db    942 ALKEPTCYLQELELVDCQLTQNCCEDL                                  968
```

```
RESULT 13
PYA6_HUMAN
ID   PYA6_HUMAN     STANDARD;      PRT;   1033 AA.
AC   P59045; Q8NBF5;
DT   28-FEB-2003 (Rel. 41, Created)
DT   28-FEB-2003 (Rel. 41, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   PYRIN-containing APAF1-like protein 6.
GN   PYPAF6.
OS   Homo sapiens (Human).
OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A. (ISOFORM 1)
RX   MEDLINE=22162427; PubMed=12019269;
RA   Wang L., Manji G.A., Grenier J.M., Al-Garawi A., Merriam S.,
RA   Lora J.M., Geddes B.J., Briskin M., DiStefano P.S., Bertin J.;
RT   "PYPAF7, a novel PYRIN-containing Apaf1-like protein that regulates
RT   activation of NF-kappa B and caspase-1-dependent cytokine
RT   processing.";
RL   J. Biol. Chem. 277:29874-29980(2002).
RN   [2]
RP   SEQUENCE FROM N.A. (ISOFORM 1).
RC   TISSUE=Lymph;
RX   MEDLINE=22388257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
```

```
Db  277  :   :         :        :         ||     |  | ||       :    ||    | |
         SRPTRGNVKTFLKEVDCCTTLQLSNGKREIYFNSFFKDRQRASAALQLVHEDSILVGLC  336

Qy  120  FVPFVCMICTVLRQQZELGRDLSRTSKTTSVYLFITSVLSS-APVADGPR-QGDLRN  178

Db  337  RVAILCWITCTVLKRQMDKGRDFQLCCQTPTDLHAHFLADALTSEAGLTANQYHLGLLKR  396

Qy  179  LCRLAREGVLGRRAQFAEKELEQELRGSKVQTLFLSKKELPGVLETEVTYQFIQSFQE  238
         ||  | |  | :: : ::::        ::|         ||::|  |  |  ::| ::
Db  239  LCLLAAGGLFLSTLNFSGEDLRCVGFTEADVSVLQAANILPSNTHKD-RYKFIHLNVQE  455

Qy  239  FLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFG  298

Db  456  FCTAIAFLM--AVPNVLIPSGSREYKEKRQYSDFNQVFTIFGLLMANRRKILETSFG  512

Qy  299  CMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDTEPEEEGEEPNYPLELLYC  358

Db  513  YQLP--MVDSFKWYS---------VGYMKHL---------DRDPEKLTHEMPLFYC  548

Qy  359  LYETQEDAFVRQALCRFPELA-LQRVRFCRMDVAVLSYCVRCCPAGAQALRLISCRIVAA  416
         ||:|:|::|:|:  | | ::|   :|| |:
Db  549  LYENREEFVKTIVDALMEVTVYLQSDK---DMWVSLYCLDYCCHLRTLKLSVQRIFQN  604

Qy  417  QEKKKKSLGKRLQA-                          SLGGGSSQGGTKQLP-  445

Db  605  KEPLIRPTASQMKSLVYWREICSLFYTMESLRELHIFDNDLNGISERILSKALEHSSCKL  664

Qy  446  --                                    ASLLHPLFQAMTDPLCHLS  464
                                               ||||||  :: :|||:::
Db  665  RTLKLSYVSTASGFEDLLKALARNRSLTYLSINCTSISLNMFSLLHDI---LHEPTCQIS  721

Qy  465  SLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTV-  523

Db  722  HSLMKCDLRASECEEIASLLISGGSLRKLTLSSNPLRSDGMNILCDALLHPNCTLISLV  781

Qy  524  --------RVQLPDPQ-RGLQYLVGMLRQ----------------SPALTTLDLSG  554
                 ||:|:||   :|  :|  |:||               :||: |||||
Db  782  LVFCCLTENCCSALGRVLLFSPTLRQLDLCVNRLKNYGLVHVTFPLLFPTCQLELHLSG  841

Qy  555  COLPAPMVTY-                      LCAVLQHQGGLQTL-----  579

Db  842  CFFSSDICQYIAIVIATNEKLRSLEIGSNKEDAGMQLLCGGLRHPNCMLVNIGLEEECML  901

Qy  580  ------SLASVELSEQSLQELQAVKRAKPDLVITHPALDGHPQPPKELIS  623
                |:||: :|  :|: |        ||:
Db  902  TSACCRSLASVLTTNKTLERL-----------NLLQNHLGNDGVAKLLESLIS  943
```

Query Match             14.3%;  Score 461;  DB 1;  Length 1033;
Best Local Similarity   22.6%;  Fred. No. 6.5e-26;
Matches  174;  Conservative 119;  Mismatches 287;  Indels 190;  Gaps  18;

```
Qy    1  MLAQPQRLFILDGADELP-ALGGPEAAPCTDPFEAASGARVLGGLSKALJPTALLVT  59
            :|:||||||: ||||||  ||| | ||| |:    ||      :|    : ||
Db  217  ILSDPKKLLFLIEDLDNIRFELNVNESALCSNSTQKVPIVLVSLLKRMAPGCWFLIS  276

Qy   60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKYFYKFFRDERRAEBRAYRFVKENETLFALC  119
```

```
RX   MEDLINE=22158257; PubMed=12477932;
RA   Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
RA   Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
RA   Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
RA   Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
RA   Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
RA   Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
RA   Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
RA   Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
RA   Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
RA   Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
RA   Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
RA   Fahey J., Helton E., Ketteman M., Madan A., Rodrigues S., Sanchez A.,
RA   Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
RA   Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
RA   Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
RA   Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
RA   Schnerch A., Schein J.E., Jones S.JM., Marra M.A.;
RT   "Generation and initial analysis of more than 15,000 full-length
RT   human and mouse cDNA sequences.";
RL   Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
CC   -!- FUNCTION: Induces NF-kappaB via RICK (CARDIAK, RIP2) and IKK-
CC       gamma. Confers responsiveness to intracellular bacterial
CC       lipopolysaccharides (LPS) (By similarity).
CC   -!- SUBUNIT: Binds to RICK by CARD-CARD interaction (By similarity).
CC   -!- SUBCELLULAR LOCATION: Cytoplasmic (By similarity).
CC   -!- ALTERNATIVE PRODUCTS:
CC       Event=Alternative splicing; Named isoforms=2;
CC       Name=1;
CC         IsoId=Q8K3Z0-1; Sequence=Displayed;
CC       Name=2;
CC         IsoId=Q8K3Z0-2; Sequence=VSP_007066, VSP_007070;
CC         Note=No experimental confirmation available;
CC   -!- SIMILARITY: Contains 2 CARD domains.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- SIMILARITY: Contains 10 leucine-rich (LRR) repeats.
CC   --------------------------------------------------------------------
CC   This SWISS-PROT entry is copyright. It is produced through a collaboration
CC   between the Swiss Institute of Bioinformatics and the EMBL outstation
CC   the European Bioinformatics Institute. There are no restrictions on its
CC   use by non-profit institutions as long as its content is in no way
CC   modified and this statement is not removed. Usage by and for commercial
CC   entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC   or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------
DR   EMBL; AF520774; AAM76073.1; -.
DR   EMBL; BC044774; AAH44774.1; ALT_INIT.
DR   MGD; MGI:2429397; Card15.
DR   InterPro; IPR001315; CARD.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   Pfam; PF00560; LRR; 2.
DR   PROSITE; PS50209; CARD; 2.
DR   PROSITE; PS50837; NACHT; 1.
KW   ATP-binding; Repeat; Leucine-rich repeat; Polymorphism;
KW   Alternative splicing.
FT   DOMAIN        6    104       CARD 1.
FT   DOMAIN      106    200       CARD 2.
FT   DOMAIN      273    600       NACHT.
FT   REPEAT      685    709       LRR 1.
FT   REPEAT      726    749       LRR 2.
FT   REPEAT      766    792       LRR 3.
FT   REPEAT      794    817       LRR 4.
FT   REPEAT      822    845       LRR 5.
FT   REPEAT      850    873       LRR 6.
FT   REPEAT      906    929       LRR 7.
FT   REPEAT      934    962       LRR 8.
FT   REPEAT      963    985       LRR 9.
FT   REPEAT     1005   1019       LRR 10.
FT   VARSPLIC      1      7       Missing (in isoform 2).
FT                                /FTId=VSP_007069.
FT   VARSPLIC    195    195       E -> EGYSLCRSRCDRGFTICLFCIL (in isoform
FT                                2).
FT   VARIANT     212    212       T -> A (in strain NMRI).
FT                                /FTId=VSP_007070.
FT   VARIANT     240    240       Q -> R (in strain NMRI).
FT   VARIANT     422    422       L -> C (in strain NMRI).
FT   VARIANT     485    485       G -> V (in strain NMRI).
FT   VARIANT     603    603       V -> A (in strain NMRI).
FT   VARIANT     675    675       V -> I (in strain NMRI).
FT   VARIANT     925    925       E -> Q (in strain NMRI).
SQ   SEQUENCE   1020 AA;  113561 MW;  25504905ECF70FBB CRC64;

Query Match             9.5%;  Score 305;  DB 1;  Length 1020;
Best Local Similarity   23.5%;  Pred. No. 1.6e-14;
Matches  157;  Conservative 111;  Mismatches 277;  Indels 124;  Gaps  28;

Qy          1 MLAQPQRLLFILDGADELPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVT   60
                                                                    : :::
Db        347 LLDHPDRVLLTFDGLDEFKFRPTDRERHCS-PIDPTSVCTLLFNLLQGVLLKKACKVLTS  405

Qy         61 RAAAPGRLQGRLCSPQCAEVRGFSDKDKKYFYKFFRDERRAERAYRFVKENE'LFALCF  120
                  :::  :|  : :|:|:||:  |||:|  |  ::|::|  ||:  | :   :|
Db        406 RPDAVSALLRKFVRTE-LQLKGFSEEGIQLYLRKHHREPGVADRLIQLIQATSALHGLCH  464

Qy        121 VPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVL-----SSAPVADGPRL--  172
                   :|:|::| : :: : :|:  : :::  : :  :       ::   :: : :::
Db        465 LPVFSWMVSRCHRELLLQNRGFPTTS--TDMYLILQHFLLHASPPDSPLGLGSPGLLQ  521

Qy        173 -QGDLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFL--SKKELPGV'ETEVT  228
                :: |  | :|: :   :::  :  : :  ::    :||: : |    ::: :::
Db        522 SRLSTLLHLGHLALRGLAMSCYVFSAQQLQAAQVDSDISLGFLVRAQSSVPG---SKAP  578

Qy        229 YQFIDQSFQEFLAA-----------------LSYLLEDGGV--------PRTAAGG  259
               |:: ::||::::                     :||:::      ::    :
Db        579 LEFLHITFQCFFAAFYLAVSADTSVASLKHLFSCGRIGSSLLGRLLPNLCIQGSRVKKGS  638

Qy        260 VGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGQGCG  319
                 ::  : |:   ::   | :::: :::  ::   :    ::  ::
Db        639 EAALLQ-KAEPH-NLQITAAFLAGLLS-QHRDLLA-ACQVSERV---LLQRQARARS  689

Qy        320 CPGVAPEVTEGAKGLEDTEPEEEEGEEPNYPLELYCLYETQEDAPVRQALCRFPELA  379
               | |:|| | |:    : : ::::| ||| : :   |: :: :|  :| : |  |:
Db        690 C--LAHSLREHFHSIPPAVPGETKSMHAMPGF-IWLIRSLYEMOEEQLAQEAVRRL-DIG  745

Qy        380 LQRVRFCRM---DVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLG---KRLQASL  432
                ::| |:       ||::  ::: :  | : : || :::   :||:  :     :: ::
Db        746 HLKLTFCRVGPAECAALAFVLQHLQRPVALQL-------------DYNSVGDVGVEOLRPCL  794

Qy        433 G------------GGSSQGTKQLPASLLHPLFQAWTDPLC-HLSSLTLSHCKLPAVCRDL  481
                             : : :| ::|:::|:| :  :: : |       : |:: ::
Db        795 GVCTALYLRENNISDRGARTLVECAL-------------REQLQKIALFNNKLTDACACSM  843

Qy        482 SEALRAAPALTELGLLHNRLSEAGLRMLSEGLA----------WPQCRVQTVRVQLPDPQ  531
               ::  :|  |: | |:::|:|    | :||: ||            |  | :::  |:: :
Db        844 AKLLAHKQNFSLRVGNNHITAAGEVLAQGLKSNTLKFLGFWGN-----SVGD--  893

Qy        532 RGLQLYVGMLRQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQSL  591
               :| :  : :|:|   : ::: :  : :|:   ::   :|:     :   :::   ::: :
Db        894 KGTQALAEVVADHQNLKWLSLVGNNI-GSMGAEALALMLEKNKSLEELCLEENHICDEGV  952

Qy        592 QEL-QAVKR  599
               : :    |:
Db        953 YSLAEGLKR  961


RESULT 15
C2TA_HUMAN
ID   C2TA_HUMAN     STANDARD;     PRT;   1130 AA.
AC   P33076;
DT   01-OCT-1993 (Rel. 27, Created)
DT   01-OCT-1993 (Rel. 27, Last sequence update)
DT   15-SEP-2003 (Rel. 42, Last annotation update)
DE   MHC class II transactivator (CIITA).
GN   MHC2TA or CIITA.
OS   Homo sapiens (Human).
```

OC   Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC   Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX   NCBI_TaxID=9606;
RN   [1]
RP   SEQUENCE FROM N.A., VARIANTS BLS II LYS-120 DEL ILE-GLU INS AND
RP   940-THR--ALA-963 DEL, AND VARIANT GLY-500.
RX   MEDLINE=94006536; PubMed=8402893;
RA   Steimle V., Otten L.A., Zufferey M., Mach B.;
RT   "Complementation cloning of an MHC class II transactivator mutated in
RT   hereditary MHC class II deficiency (or bare lymphocyte syndrome).";
RL   Cell 75:135-146(1993).
RN   [2]
RP   SEQUENCE FROM N.A.
RX   MEDLINE=95269136; PubMed=7749984;
RA   Riley J.L., Westerheide S.D., Price J.A., Brown J.A., Boss J.M.;
RT   "Activation of class II MHC genes requires both the X box region and
RT   the class II transactivator (CIITA).";
RL   Immunity 2:533-543(1995).
RN   [3]
RP   VARIANT BLS II GROUP A SER-962.
RX   MEDLINE=99432007; PubMed=10501838;
RA   Quan V., Towey M., Sacks S., Kelly A.P.;
RT   "Absence of MHC class II gene expression in a patient with a single
RT   amino acid substitution in the class II transactivator protein
RT   CIITA.";
RL   Immunogenetics 49:957-963(1999).
RN   [4]
RP   VARIANT BLS II GROUP A PRO-469.
RX   MEDLINE=21359573; PubMed=11466404;
RA   Wiszniewski W., Fondaneche M.-C., Le Deist F., Kanariou M., Seiz F.,
RA   Brousse N., Steimle V., Barbieri G., Alcaide-Loridan C., Charron D.,
RA   Fischer A., Lisowska-Grospierre B.;
RT   "Mutation in the class II trans-activator leading to a mild
RT   immunodeficiency.";
RL   J. Immunol. 167:1787-1794(2001).
RN   [5]
RP   ERRATUM.
RA   Wiszniewski W., Fondaneche M.-C., Le Deist F., Kanariou M., Seiz F.,
RA   Brousse N., Steimle V., Barbieri G., Alcaide-Loridan C., Charron D.,
RA   Fischer A., Lisowska-Grospierre B.;
RL   J. Immunol. 169:607-607(2002)
RN   [6]
RP   VARIANTS BLS II GROUP A 964-LEU--ASP-991 DEL AND ILE-1027 DEL.
RX   MEDLINE=21850498; PubMed=11862382;
RA   Dziembowska M., Fondaneche M.-C., Vedrenne J., Barbieri G.,
RA   Wiszniewski W., Picard C., Cant A.J., Steimle V., Charron D.,
RA   Alca-Loridan C., Fischer A., Lisowska-Grospierre B.;
RT   "Three novel mutations of the CIITA gene in MHC class II-deficient
RT   patients with a severe immunodeficiency.";
RL   Immunogenetics 53:822-829(2002).
CC   -!- FUNCTION: ESSENTIAL FOR TRANSCRIPTIONAL ACTIVITY OF THE HLA CLASS
CC       II PROMOTER. ACTIVATION IS VIA THE PROXIMAL PROMOTER. NO DNA
CC       BINDING OF IN VITRO TRANSLATED CIITA WAS DETECTED. MAY ACT IN A
CC       COACTIVATOR-LIKE FASHION THROUGH PROTEIN-PROTEIN INTERACTIONS BY
CC       CONTACTING FACTORS BINDING TO THE PROXIMAL MHC CLASS II PROMOTER,
CC       TO ELEMENTS OF THE TRANSCRIPTION MACHINERY, OR BOTH. ALTERNATIVELY
CC       IT MAY ACTIVATE HLA CLASS II TRANSCRIPTION BY MODIFYING PROTEINS
CC       THAT BIND TO THE MHC CLASS II PROMOTER.
CC   -!- SUBCELLULAR LOCATION: Nuclear.
CC   -!- DISEASE: Defects in MHC2TA are a cause of bare lymphocyte syndrome
CC       type II (BLS II) [MIM:209920]; also known as hereditary MHC class
CC       II deficiency or HLA class II-deficient combined immunodeficiency.
CC       BLS II is a form of severe combined immunodeficiency.
CC       (SCID) characterized by a profound defect in constitutive and
CC       interferon-gamma induced MHC II expression, an absence of cellular
CC       and humoral T-cell response to antigen challenge. The
CC       hypogammaglobulinemia and impaired antibody production. The
CC       consequence include extreme susceptibility to viral, bacterial and
CC       fungal infections. MHC2TA is linked with BLS II complementation
CC       group A.
CC   -!- SIMILARITY: Contains a leucine-rich (LRR) repeats.
CC   -!- SIMILARITY: Contains 1 NACHT domain.
CC   -!- DATABASE: NAME=Atlas Genet. Cytogenet. Oncol. Haematol.;

CC       WWW="http://www.infobiogen.fr/services/chromcancer/Genes/MHC2TAID260.html".
CC   -!- This SWISS-PROT entry is copyright. It is produced through a collaboration
CC       between the Swiss Institute of Bioinformatics and the EMBL outstation
CC       the European Bioinformatics Institute. There are no restrictions on its
CC       use by non-profit institutions as long as its content is in no way
CC       modified and this statement is not removed. Usage by and for commercial
CC       entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC       or send an email to license@isb-sib.ch).
CC   --------------------------------------------------------------------------
DR   EMBL; X74301; CAA52354.1; -.
DR   EMBL; U18259; AAA88861.1; -.
DR   PIR; A48843; A48843.
DR   Genew; HGNC:7067; MHC2TA.
DR   MIM; 600005; -.
DR   MIM; 209920; -.
DR   GO; GO:0003677; F:DNA binding activity; TAS.
DR   GO; GO:0003702; F:RNA polymerase II transcription factor acti... .; TAS.
DR   GO; GO:0003713; F:transcription co-activator activity; TAS.
DR   GO; GO:0006955; P:immune response; TAS.
DR   GO; GO:0009596; P:perception of pest/pathogen/parasite; TAS.
DR   InterPro; IPR000767; Disease_resist.
DR   InterPro; IPR001611; LRR.
DR   InterPro; IPR007091; LRR_RNinh.
DR   InterPro; IPR007111; NACHT_NTPase.
DR   Pfam; PF05560; LRR; 1.
DR   PRINTS; PR00364; DISEASERSIST.
DR   PROSITE; PS50837; NACHT; 1.
KW   Transcription regulation; Activator; Nuclear protein; ATP-binding;
KW   Leucine-rich repeat; Repeat; Disease mutation; SCID; Polymorphism.
FT   REPEAT         98    121       LRR 1.
FT   REPEAT        214    238       LRR 2.
FT   DOMAIN        414    724       NACHT.
FT   REPEAT       1014   1037       LRR 3.
FT   REPEAT       1071   1094       LRR 4.
FT   DOMAIN         52    137       ASP/GLU-RICH (ACIDIC).
FT   NP_BIND       420    427       ATP (POTENTIAL).
FT   VARIANT       120    120       K -> IE (in BLS II group A).
FT                                  /FTId=VAR_005127.
FT   VARIANT       469    469       L -> ? (in BLS II group A; mild
FT                                  immunodeficiency. Has residual MHC class
FT                                  II trans activation activity).
FT                                  /FTId=VAR_015551
FT   VARIANT       500    500       A -> G (in dbSNP:4774).
FT                                  /FTId=VAR_005128.
FT   VARIANT       658    658       A -> G (in dbSNP:2229319).
FT                                  /FTId=VAR_015552.
FT   VARIANT       940    963       Missing (in BLS II group A).
FT                                  /FTId=VAR_005129.
FT   VARIANT       962    962       F -> S (in BLS II group A).
FT                                  /FTId=VAR_015553.
FT   VARIANT       964    991       Missing (in BLS II group A).
FT                                  /FTId=VAR_015554.
FT   VARIANT      1027   1027       Missing (in BLS II group A).
FT                                  /FTId=VAR_015555.
SQ   SEQUENCE   1130 AA;  123456 MW;  DC4D0818O2987E06 CRC64;

Query Match             9.5%;  Score 304.5;  DB 1;  Length 1130;
Best Local Similarity   25.5%;  Pred. No. 2e-14;
Matches  174;  Conservative   80;  Mismatches  262;  Indels  167;  Gaps   29;

Qy       1 MLAQPQRLFILDGADELPALGGPEAPC-TDPFEAASGARVLGGLLSKALLPTALLLVT   59
Db     487 ILKRPDRVLLILDAFEELEAQDGFLHSTCGPAPAEPCSLRGLLAGLFQKKLLRGCTLLLT  546

Qy      60 TRAAAPGRLQGRLCSPQCA-EVRGFSDKDKKKYFYKFPRDERRAE--RAYRFVKENETL  115
                                ||||        ::|:                  :|:
Db     547 AR?R--GRLVQSLSKADALFELSGFSMEQAQAYVMRYFESSGMTEHQDRALTLRDRPLL  604

Qy     116 FALCFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYL-LFITSVLSSAPVADGPRLQG  174
             ::| .:| |||:  |||| :|||:  |:| ::| :::|:|:| :|:  |||
Db     605 LSHSHSPTLCRAVCQLSEALLELGED-AKLPSTLTGLYVGLLGRAALDSPP-------G  655

```
Qy   175  DLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFLSK---KELPGVLETEVTY-Q  230
               ::|:|||  |||| : |||       |  : |:|     ||:|         ||:|
Db   656  ALAELAKLAWE--LGRRHQ---STLQEDGFPSADVRTWAMAKGLVQHPPRAAESELAFPS  710

Qy   231  FIDQSFQEFL-AALSYLLEDSGVPRTAAGGVGTLLRGDAQPHSH-LVLTRFLFGLLSAE  288
               ::|: :  | ::||| :||| |   :  |::   :| |     | ||:| :||
Db   711  FLLQCFLGALWLALSGEIKDKELPGYLA----LTPRKKRPYDNMLEGVPRFLAGLLFQP  765

Qy   289  RMRDIERHFGCMVSERV--KQEAL-RWVQGQGCPGVAPEVTEGAKGLEDTSEPEEEE    345
               : | ||| : :|: |  |:|| |  |       : :|                   |
Db   766  PARCLGALLGPSAAASVDRKQKVLARYLKRKLQPG--------TLRARQL------    806

Qy   346  GEEPNYPLELLYCLYETQEDAFVRQALCRFP-ELALQRVRFCRMDVAV-SYCVRC----  399
               ||:| |:: :|: :: ||| |:       :     |       |:|
Db   807  ------LELLHCAHEAEEASIWQHVVQELPGRLSFLGTRLTPPDAHVLGKALEAAGQDF  859

Qy   400  ---------CPAG-QALRLISC------------------------RLVAAQEK----  419
                        |: || : |:|:|                          ||
Db   860  SLDLRSTGICPSGLGSLVGLSCVTRFRAALSDTVALWESLRQHGETKLLQAAEEKFTIEP  919

Qy   420  -KKKS---IGKRLQASLGGGSQGTTKQLPA-------------SLLHPLFQAMTDP---  459
            |:|:    |||:|||| :  :   | |:|                :|:|| ||:|:
Db   920  FKAKSLKDVEDLGKLVQTQRTRSSEDTAGELPAVRDLKKLEFALGPVSGPQAFPKLVRI  979

Qy   460  ----LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGL  513
                 :|  |  : | :|    |  :|  | :|       |: |   |     |:  :|
Db   980  LTAFSSLLQHLDLDALSENKISDEGVSQLSATFPQLKSLETLNLSQNNITDLGAYKLAEAL  1039

Qy   514  AWPQCRVQTVRVQ-----------LPD------------PQRGLQYLVGNLR  542
             |                     |:                 |       : |
Db   1040  --PSLAASLLRLSLYNKCICDVGAESLARVLPDMVSLRVMDVQYNKFTAAGAQQLAASLR  1097

Qy   543  QSPALTTLDLSGCQLPAPMVTYL  565
                ::   |  : :   | :|
Db   1098  RCPHVETLAMWTFTIPFSVQEHL  1120
```

Search completed: October  2, 2003, 17:49:24
Job time : 31 secs

Copyright GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on: October 2, 2003, 17:47:48 ; Search time 43 Seconds
(without alignments)
1397.801 Million cell updates/sec

Title: US-10-029-347-2
Perfect score: 3218
Sequence: 1 MLAQPQRLFLFILDGADELPA.........ITHPALDGHPQPPKELISTP 625

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 283308 seqs, 96168682 residues

Total number of hits satisfying chosen parameters: 283308

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_76:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Match Length | DB | DB ID | Description |
|---|---|---|---|---|---|---|
| 1 | 1578 | 49.0 | 483 | 2 | S27880 | Nasopressin recept |
| 2 | 762 | 23.7 | 1192 | 2 | T17255 | hypothetical prote |
| 3 | 468.5 | 14.6 | 1111 | 2 | A59000 | mater protein (imp |
| 4 | 304.5 | 9.5 | 1130 | 2 | A48843 | MHC class II trans |
| 5 | 216 | 6.7 | 456 | 2 | A31857 | ribonuclease inhib |
| 6 | 212 | 6.6 | 461 | 2 | A31858 | ribonuclease-angio |
| 7 | 209.5 | 6.5 | 456 | 2 | S20597 | ribonuclease inhib |
| 8 | 170 | 5.3 | 1004 | 2 | T31665 | hypothetical prote |
| 9 | 130 | 4.0 | 914 | 2 | T17233 | hypothetical prote |
| 10 | 116 | 3.6 | 534 | 2 | S75472 | GTP-binding protei |
| 11 | 114.5 | 3.6 | 877 | 2 | T35861 | probable large sec |
| 12 | 114 | 3.5 | 1075 | 2 | T31668 | hypothetical prote |
| 13 | 113.5 | 3.5 | 681 | 2 | AF0697 | probable type III |
| 14 | 112.5 | 3.5 | 959 | 2 | C82392 | oxidoreductase/iro |
| 15 | 112.5 | 3.5 | 1236 | 2 | T03465 | probable exonuclea |
| 16 | 111 | 3.4 | 2274 | 2 | T30258 | adenomatous polypo |
| 17 | 110 | 3.4 | 2059 | 2 | T13858 | probable DNA-direc |
| 18 | 108 | 3.4 | 642 | 2 | E97332 | cobalamin biosynth |
| 19 | 107.5 | 3.3 | 780 | 1 | WXBEH8 | infected cell prot |
| 20 | 106 | 3.3 | 906 | 2 | G83156 | probable transcrip |
| 21 | 106 | 3.3 | 1753 | 2 | T00350 | hypothetical prote |
| 22 | 105 | 3.3 | 506 | 2 | A45841 | T-complex-associat |
| 23 | 104.5 | 3.2 | 343 | 2 | AE2045 | hypothetical prote |
| 24 | 104.5 | 3.2 | 2176 | 2 | T13806 | toucan gene protei |
| 25 | 104 | 3.2 | 857 | 2 | JC4169 | phosphoenolpyruvat |
| 26 | 103 | 3.2 | 1189 | 1 | GNMVM7 | HIV-1 retropepsin |
| 27 | 102.5 | 3.2 | 383 | 1 | DEHU | Ig delta chain C r |
| 28 | 102.5 | 3.2 | 509 | 2 | S17597 | Ig delta chain (WI |
| 29 | 102 | 3.2 | 507 | 1 | FWCNBA | beta-globulin A pr |

| 30 | 102 | 3.2 | 1148 | 2 | T13347 | Cnn protein - frui |
| 31 | 101.5 | 3.2 | 535 | 2 | T48102 | RAN GTPase activat |
| 32 | 101.5 | 3.2 | 736 | 2 | T03849 | Fas-binding protei |
| 33 | 101 | 3.1 | 645 | 2 | H85034 | hypothetical prote |
| 34 | 101 | 3.1 | 852 | 2 | AD0050 | probable phosphoen |
| 35 | 100.5 | 3.1 | 772 | 2 | T52402 | hypothetical prote |
| 36 | 100.5 | 3.1 | 1010 | 2 | T36383 | probable large ATP |
| 37 | 100 | 3.1 | 638 | 2 | A36929 | virulence regulato |
| 38 | 100 | 3.1 | 730 | 2 | T33041 | hypothetical prote |
| 39 | 99.5 | 3.1 | 456 | 2 | T48291 | hypothetical prote |
| 40 | 99.5 | 3.1 | 892 | 2 | T29420 | probable transcrip |
| 41 | 99.5 | 3.1 | 1079 | 2 | F82447 | pyruvate-flavoredo |
| 42 | 99 | 3.1 | 2509 | 2 | G01880 | fatty-acid synthas |
| 43 | 98.5 | 3.1 | 327 | 2 | AH3363 | alkanal monooxygen |
| 44 | 98.5 | 3.1 | 740 | 2 | T03847 | Fas-binding protei |
| 45 | 98.5 | 3.1 | 873 | 2 | T16282 | hypothetical prote |

ALIGNMENTS

RESULT 1
S27880
Nasopressin receptor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 17-Apr-1993 #sequence_revision 17-Apr-1993 #text_change 05-Nov-1999
C;Accession: S27880
R;Herrera, V.L.; Ruiz-Opazo, N.
submitted to the EMBL Data Library, February 1992
A;Description: Characterization of a cDNA encoding an AII and AVP receptor isoform.
A;Reference number: S27880
A;Accession: S27880
A;Molecule type: mRNA
A;Residues: 1-483 <HER>
A;Cross-references: EMBL:M85183; NID:g202805; PIDN:AAA03623.1; PID:g202806
C;Genetics:
A;Gene: AVP

Query Match 49.0%; Score 1578; Pred. No. 2.6e-112;
Best Local Similarity 67.2%; DB 2; Length 483;
Matches 330; Conservative 48; Mismatches 97; Indels 16; Gaps 6;

Qy    136 LELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFA 195
          ::||||||||||||||||||: |||||::|: ||||: |:|||:||||||:||||| ::|:
Db      1 MELGRDLSRTSKTTTSVYLLF-TSMLKSAG-TNGPRVQGELRMLCRLAREGILKHQAQFS 59

Qy    196 EKELEQELR--GSKVQTLF:SKKELPGVLETEV:YQFIDQSFQEFLAALSYLLEDGGVP 253
          |||||||||  :|||||||::|||||||||||||:||||||||||||||||||::|||
Db     60 EKDLERLKLKLQGSQVQTMFLSKKELPGVLETVVTVQFIDQSFQEFLAALSYLLEGAP 119

Qy    254 RTAAGGVGTLLRGDAQPHSHLVLTTRFLFG:LSAERMRDIERHFGCMVSERVKQEALRWV 313
          ::|||::|:|||:|||:|||:||||||||||| |:|:||::||:||:|||::||||||||
Db    120 GNSAGSVQMLNSDAGLRGHLALTTRFLFGLLSTERIRDIGNHFGCVVPGRVKQDTLRWV 179

Qy    314 QGQGGCPGVA---PEVTEGAKGLEDTEEPEEBEGGEEPNYP:ELJYCLYETQEDAFVRQ 370
          ||:| ||::|   | |: :|||::|::|||||||||::||||:| |||||||||||||||
Db    180 QGQSQ--PKYATVGAEKKDELKDEAAEEEEEEEELNFGLELLYCLYETQEDDFVRQ 237

Qy    371 ALCRFPELALQRVRFCRMDVAVLSYCRCCPAGQALRLISCRLVAQE--KKKKSLGKRL 428
          |||::||||||||:|:||||::||||||::|||||||:||||:|||  |||||||||||
Db    238 ALSSLPEMVLERVRLITRMDLEVLSYCVQCCPDGQALRLVSCGLVAAKEKKKKKSFMNRL 297

Qy    429 QASLGGGSSOGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAA 488
          ||||||||||||||||||||:||||||||||:|||||||||||||||||||||||||||
Db    298 K-----GSQSTGKQPPASLLRPLCEAMITQQCGLSILTLSHCKLPDAVCRDLSEALKVA 351

Qy    489 PALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQSPALT 548
          |:|||:||||:||:||||||:|||||||:|:|:|||||||||||:|:|||:|||||||||
Db    352 PSLRELGLLQNRLTEAGLRLLSQGLAWPKCKCKVQTLRIQMPGLQEVYHYLVIVLQQSPVLT 411

Qy    549 TLDLSGCQLPAPMV:YLCAVLQHOGCGLQTLSLASVELSEGSLQELQAVKRAKPDLVITH 608
          ||||||||||||||::|||||||||||||:|:|||:|||||||:| |||:||||||||||:
Db    412 TLDLSGCQLPGTVVEPLCSALKHPKCGLKTLSLTSVELTENPLRELQAVKTLKPDLAIIH 471

Qy  609  PALDGHPQPPK 619
         :|||||||
Db  472  SKLGTHPQPLK 482

RESULT 2
T17255
hypothetical protein DKFzp586O1822.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17255
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18722
A;Accession: T17255
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-192 <KCE>
A;Cross-references: EMBL:AL117470
A;Experimental source: adult uterus; clone DKFzp586O1822
C;Genetics:
A;Note: DKFZp586O:822.1

Query Match         23.7%;  Score 762;  DB 2;  Length 1192;
Best Local Similarity  33.2%;  Pred. No. 1.1e-49;
Matches 220;  Conservative 103;  Mismatches 207;  Indels 132;  Gaps 22;

Qy    1  MLAQPQRLFLFLDGADELP--ALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLV 58
          ||||||||||| :|||||   ||||||||||||||||||||||| |||| :|||||||||
Db  117  ILSRPERLLFIIDGVDE-PGWVLQEBSSELCLHWSQPPADALLGSLLGKTILFEASFLI 175

Qy   59  TRRAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 118
          :||||||| |||:|||| ||||||||||||||||||||||||||||||||||||||| |
Db  176  TARTTALQNLIPSLEQARWVEVLGFSESSRKEYFYRYFTDERQAIRAFRLVKSMKELWAL 235

Qy  119  CFVPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRN 176
          || ||||||||||| ||||||| ||||||||||||| : :||||||||||||| :|||
Db  236  CLVPWVSWLACCLMQQMKREKJLTSKTTTLCLHYLAQALCAQPL--GFC---LRD 289

Qy  177  RNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFSKKEPGVLETEVTYQFIDQSF 236
          :|||||||||| ||||||||||||||||| ||||||||:||||||||||||||||||
Db  290  LCSLAAEGIWQKKTLFSPDEIRKHGLDGAIIST-FLKM---GILQEHPIPLSYSFIHLC 344

Qy  237  QEFLAALSYLLEDGGVPRTAAGGV-------------GTLLRGDAQPHSHLVLTTRFLFG 283
          ||| ||||| |||||||||||||            |:|||||||||||:|||||||||
Db  236  FQEFLAALSYLLEDGGVPRTAAGGV-GLLRGDACPSHLVL---------EKGRG 277

Qy  284  LLSAERXRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEETEEPEEE 343
          ||||||||:||:||||||||||||||||| : ||||||:|||||||||||||||
Db  345  FQEFFAAMSYYLED----------KHSNCIIDLEKTLEAYGIHGLFGAST 389

Qy  344  LMNKDILKTLEVLFEYPVIPTVEQKLQHWVSLIAQQVNGTSPMDT--------------- 403
          |:|||||||||||||||||||||||||||||||||||||||||
Db  278  TRFLFGLLSAERMRDIER:FGCVVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEDT 337

Qy  404  EEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAG 403
          |||||||||||||||:||||||||||||||||||||||||-||||||||:||||
Db  390  TRFLLGLLSDEGEREMENIFHCRLSQ--GRNLMQWV---------------PSLQLL 429

Qy  595  --LDAFYCLFESQDEEFVGGALKRFQEVWLLINQ--KMDLKVSSYCLKHCQNL 643
           ||||||||||||||||||||||||||||||||   ||||||||||||||||
Db  338  EEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVFC--RMDVAVLS 394

Qy  644  KAIRVDIRDLLSVDNTLELCPVVTVQETQCKPLLMEWWNGNFCSVLGSLRNLKELDLGDSI 703
          :||||:||||||||||||||||||||||||||||||||||||||||||||||||||:|||
Db  430  LQP-------HSLESLFCLYETRNKTFLTQVMAHFEEWGM------CVETDXELLVCT 474

Qy  704  LSQRAMKILCLELRNQSCRIQKLTFKSAEVVSGLKHLWKLLFSNQNLKYLNLGNTPMKDD 763
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  395  YCVRCCPAGQALRLISCR----------LVAAQEKKKKSLGKRJQ 429

Qy  442  --KQLPASLLHP- 451
          | |||||||||
Db  475  FCIKFSRHVKKJQLIEGRQERSTWSPTMVVLFRWVPTDAYWQILFSVLKVTRNJ-KELD 533

Qy  764  DMKLACEALKHPKCSVETLRLDSCELTIIGYEMISTLLISTRLKCLSLAKNRVGVKSMI 823
          ||| |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  436  ---------------SSQ----------GTT-- 441

Qy  452  -LFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLS 510
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  704  LSQRAMKILCLELRNQSCRIQKLTFKSAEVVSGLKHLWKLLFSNQNLKYLNLGNTPMKDD 763

Qy  824  SLGNALSSSMCLLQKLILDNCGLTPASCHLLVSALFSNQNLTHLCLSNNSLGTEGVQQLC 883
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  442  --KQLPASLLHP 451

Qy  511  EGLAWPQCRVQTV--RVQLPDDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCA 567
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db  764  DMKLACEALKHPKCSVETLRLDSCELTIIGYEMISTLLISTRLKCLSLAKNRVGVKSMI 823

Qy  884  QFLRNPECALQRILLNHCNIVDDAYG--FLAMRLANNTKLTHLSIMNPVGDGAXKLLCE 941
          ||||||||||||||||||||||||||  |||||||||||||||||||||||||||||||
Db  605  VI 606

Db  701  LI 702

RESULT 3
A59000
mater protein [imported] - mouse
C;Species: Mus musculus (house mouse)
C;Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C;Accession: A59000
R;Tong, Z.B.; Nelson, L.M.
Endocrinology 140, 3720-3726, 1999
A;Title: A mouse gene encoding an oocyte antigen associated with autoimmune premature
A;Reference number: A59000; MUID:99360614; PMID:10433232
A;Accession: A59000
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-1111 <KUR>
A;Cross-references: GB:AF074018; NID:g5802697, PIDN:AAD51762.1; PID:g5802698
C;Genetics:
A;Gene: Mater

Query Match         14.6%;  Score 468.5;  DB 2;  Length 1111;
Best Local Similarity  23.8%;  Pred. No. 2.3e-27;
Matches 178;  Conservative 108;  Mismatches 270;  Indels 191;  Gaps 15;

Qy    1  MLAQPQRLFLFLDGADELPALPALGGPEAAPCTDPFEAASGARVL-GGLLSKALLPTALLLVT 59
          |||||||||||||||| : :| |||||||||||||||||||||| |||:||||||||||||:
Db  260  IMSQPERLLFVIDGLDDMSVLQHDDMTLSRDWKDEQFIYILMYSLLRKALLPQSFLIT 319

Qy   60  TRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALC 119
          |||||||||||||| |||||||||||||||||||||||||||||||||||||||||||
Db  320  TRNTGLEKLKSMVVSPLYILVEGLSASRRSQLVLENISNESDRIQVFHSLIENHQLFDQC 379

Qy  120  FVPFVCWIVCTVLRQQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGD---L 176
          | ||||||||||||||||||||||||||||||||||||||||||||||||||||   |
Db  380  QAPSVCSLVCEALQLQKKLGKRCTLPCQTLGTLVATLVFHQLTLKRPSQSALSQEEQITL 439

Qy  177  RNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTLFSKKELPGVLETEVTYQFIDQSF 236
          ||||||||||||||||||||||||||||||||||||||| ||||||||||||||||||
Db  440  VGLCMVAAEGVWTMRSVFYDDDLKNYSLKESEILALFHMNILQVGHNSEQCYVFSHLSL 499

Qy  237  QEFLAALSYLLEDGGVPRTAAGGV-----------GTLLRGDAQPHSHLVLTTRFLFG 283
          |||||||||| |||||||||||||          |:|||||||||||:||||||||||
Db  500  QCFFAALYVLE-----------GLEEWNQHCFIENQRSIMEVKRTDDTRLLGMKRFLFG 549

Qy  284  LLSAERXRDIERHFGCMVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGLEETEEPEEE 343
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  550  LMNKDILKTLEVLFEYPVIPTVEQKLQHWVSLIAQQVNGTSPMDT--------------- 594

Qy  344  EEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVFCRMDVAVLSYCVRCCPAG 403
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  595  --LDAFYCLFESQDEEFVGGALKRFQEVWLLINQ--KMDLKVSSYCLKHCQNL 643

Qy  404  QALR-----LIS-----CRLVAAQEKKK------------SLGKRLQASLGGG- 435
          ||||     |||     ||||||||||               |||||||||||
Db  644  KAIRVDIRDLLSVDNTLELCPVVTVQETQCKPLLMEWWNGNFCSVLGSLRNLKELDLGDSI 703

Qy  436  -----------------------------------------SSQ----------GTT 441
          |||||||||||||||||||||||||||||||||||||||| |||          |||
Db  704  LSQRAMKILCLELRNQSCRIQKLTFKSAEVVSGLKHLWKLLFSNQNLKYLNLGNTPMKDD 763

Qy  442  --KQLPASLLHP- 451
          | |||||||||
Db  764  DMKLACEALKHPKCSVETLRLDSCELTIIGYEMISTLLISTRLKCLSLAKNRVGVKSMI 823

Qy  452  -LFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLS 510
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  824  SLGNALSSSMCLLQKLILDNCGLTPASCHLLVSALFSNQNLTHLCLSNNSLGTEGVQQLC 883

Qy  511  EGLAWPQCRVQTV--RVQLPDDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMVTYLCA 567
          |||||||||||||  ||||||||||||||||||||||||||||||||||||||||||||
Db  884  QFLRNPECALQRILLNHCNIVDDAYG--FLAMRLANNTKLTHLSIMNPVGDGAXKLLCE 941

```
Qy   568 VLQHQGGLQTLSLASVELSEQSLQEL 594
Db   942 ALKEPTCYLQEILELVDCCLTQNCCEDL 968
```

RESULT 4
A48843
MHC class II transactivator - human
N;Alternate names: CIITA
C;Species: Homo sapiens (man)
C;Date: 06-Jan-1995 #sequence_revision 06-Jan-1995 #text_change 08-Oct-1999
C;Accession: A48843
R;Steimle, V.; Otten, L.A.; Zufferey, M.; Mach, B.
A;Title: Complementation cloning of an MHC class II transactivator mutated in hereditary
A;Reference number: A48843; MUID:94096536; PMID:8402893
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-1130 <STE>
A;Cross-references: GB:X74321; NID:g414112; PIDN:CAA52354.1; PID:g414113

Query Match      9.5%;      Score 304.5;  DB 2;  Length 1130;
Best Local Similarity   25.5%;  Pred. No. 7.2e-15;
Matches 174;  Conservative 80;  Mismatches 262;  Indels 167;  Gaps 29;

```
Qy     1 MLAQPQRLLFILDGADELPALGGPEEAAPC-TDPFEAASGARVLGGLLSKALLPTALLLVT 59
           :||::|   |:|   |   :|::  | : ||       :|     |  |::|   | | |
Db   487 ILKRPDRVLLILDAFEELEAQDGFLHSTCGPARAEPCSLRGLLAG-FQKKLLRGCTLLLT 546

Qy    60 TRAAAPGRLQGRLCSPQCA-EVRGFSDKDKDKKYFYKFPRDERRAE--RAYRFVKENET. 115
Db   547 ARPR--GRLVQSLSKADALFELSGFSMEQAQAYVMRYFESSGMTEHQDRALTLLRDRPLL 604

Qy   116 FALCFVPFVCWIVCTVLRQQLELGRDLSRTSKTTTSVYJ-LFITSVLSSAPVADGPRLQG 174
Db   605 LSHSHSPTLCRAVCQLSEALLELGED-AKLPSTLTGLYVGLGRAALDSPP--------G 655

Qy   175 DLRNLCRLAREGVLGRRAQFAEKELEQELRGSKVQTLFLSK--KELPGVLETEVTY-Q 230
Db   656 ALAELAKLAWE--LGRRHQ--STLQEDQFPSADVRTWAMAKGLVQHPPRAAESELAFPS 710

Qy   231 FIDQSFQEFL-AALSYLLEDGGVPRTAAGGVGT-LRGDAQFHSH-LVLTTRFLFGLLSAE 288
Db   711 FLLQCF-GALWLALSGEIKDFELPQYLA----LTFRKKRPYDNWLEGVPRFFLAGLIFQF 765

Qy   289 RMRDIERHFGCMVSERV--KQEAL-RWVQGGQGCPGVAPEVTEGAKGLEDTEEPEEEEE 345
Db   766 PARCLGALLGPSAAASVDRKQKVLARYLKRLQPG-----------TLRRAQL------ 806

Qy   346 GEEPNYPLELJCLYETQEDAFVRQA=CRFP-ELALQRVRFCRMDVAVLSYCVRC---- 399
Db   807 ---LELLHCAHEAEAGIWQHVVQELPGRLSFLGTRLTPPDAHVLGKALEAAGQDF 859

Qy   400 ------CPAG-QALRLISC-------RLVAAQEK---- 419
Db   860 SLDLRSTGICPSGLGSLVGLSCVTRFRAALSDTVALWESLRQHGETKLLQAAEEKFT-EP 919

Qy   420 -XKKS----LGKRLQASLGGGSSQGTTKQLPA------SLLHPLFQAMTDP----- 459
Db   923 FKAKSLKDVEDLGKLVQTQRTRSSSEDTAGEJPAVRDLKKLEPALGPVSGPQAFPKJVRI 979

Qy   460 -----LCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGL 513
Db   980 LTAFSS=QHLDLDALSENKIGDEGVSQLSATFPQLKSLETLNLSQNNITDLGAYKLAEAL 1035

Qy   514 AWPQCRVQTVRVQ---------LPD----------PQRGLQYLVGMLR 542
Db  1040 -PSLAASLLRLSLYNNCICDVGAESLARVLPDMVSLRVMDVQYNKFTAAGAQQLAASLR 1097

Qy   543 QSPALTTLDLSGCQLPAPMVTYL 565
```

Db  1098 RCPHVET:AMWTP:IPPSVQEHL 1120

RESULT 5
A31857
ribonuclease inhibitor, hepatic - pig
C;Species: Sus scrofa domestica (domestic pig)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 21-Jan-2000
C;Accession: A31857; A35830
R;Hofsteenge, J.; Kieffer, B.; Matthies, R.; Hemmings, B.A.; Stone, S.R.
Biochemistry 27, 8537-8544, 1988
A;Title: Amino acid sequence of the ribonuclease inhibitor from porcine liver reveals
A;Reference number: A31857; MUID:89118268; PMID:3219361
A;Accession: A31857
A;Molecule type: protein
A;Residues: 1-456 <HOF>
R;Vicentini, A.M.; Kieffer, B.; Matthies, R.; Meyhack, B.; Hemmings, B.A.; Stone, S.R.
Biochemistry 29, 8827-8834, 1990
A;Title: Protein chemical and kinetic characterization of recombinant porcine ribonucl
A;Reference number: A35830; MUID:91104783; PMID:2271559
A;Accession: A35830
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 82-456 <VIC>
A;Cross-references: GB:M58700; NID:g164638; PIDN:AAA63448.1; PID:g164639; GB:J02925
C;Superfamily: ribonuclease inhibitor; leucine-rich alpha-2-glycoprotein repeat homolo
C;Keywords: liver
F;280-304/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>

Query Match       6.7%;      Score 216;  DB 2;  Length 456;
Best Local Similarity   27.4%;  Pred. No. 1.3e-08;
Matches 77;  Conservative 45;  Mismatches 113;  Indels 46;  Gaps 7;

```
Qy   354 ELLYCLYETQ---EDAFVRQALCRFPELALQRVFECRMDVAVLSYCVRCCPAG---- 403
Db    17 ELLPLLQQYEVVRLDCGLTEHCKDIGSAL-----RANPSLTELCLRTNELGDAGVHL 70

Qy   404 --QALRLISCRLVAAQEKKKKSLGKRLQASLGGGSSQGTTKQLPA------SL 448
Db    71 VLQGLQSPTCKI----QKLSLQNCSLTEAGCGVJPSTLRSLPTLRELHLSDNPLGDAG 124

Qy   449 LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRM 509
Db   125 LRLLCEGLLDDPQCHLEKLQEYCRLTAASCEPLASVLRATRALKELTVSNNDIGEAGARV 184

Qy   509 LSEGLAWPQCRVQTVRVQLPDDPQRGL----QYLVGMLRQSPALTTLDJSGCQLPAPMVT 563
Db   185 LGQGLADSACQCLETLRLE----NCGLTPANCKDLCGIVASQASLRELDLGSNGLGDAGIA 240

Qy   564 YLCAVLQHQGCGLQTLSLASVELSEQSLQEIQAVKRAKPDL 604
Db   241 ELCPGLLSPASRLKTLWLWECDITASGCRDLCRVLQAKETL 281
```

RESULT 6
A31858
ribonuclease-angiogenin inhibitor - human
N;Alternate names: ribonuclease inhibitor, placental
C;Species: Homo sapiens (man)
C;Date: 07-Jun-1990 #sequence_revision 07-Jun-1990 #text_change 19-May-2000
C;Accession: A31858; S02012; S23933; S48636; T47188
R;Lee, F.S.; Fox, E.A.; Zhou, H.M.; Strydom, D.J.; Vallee, B.L.
Biochemistry 27, 8545-8553, 1988
A;Title: Primary structure of human placental ribonuclease inhibitor.
A;Reference number: A31858; MUID:89118269; PMID:3219362
A;Accession: A31858
A;Molecule type: mRNA
A;Residues: 1-461 <LEE>
A;Cross-references: GB:M22414; NID:g186260; PIDN:AAA59130.1; PID:g307040
R;Schneider, R.; Schneider-Scherzer, E.; Thurnher, M.; Auer, B.; Schweiger, M.
EMBO J. 7, 4151-4156, 1988
A;Title: The primary structure of human ribonuclease/angiogenin inhibitor (RAI) disclo
A;Reference number: S02012; MUID:89210799; PMID:3243277

A;Accession: S02012
A;Molecule type: mRNA
A;Residues: 1-422, 'SE',425-461 <SCH>
A;Cross-references: EMBL:X13973; NID:g35843; PIDN:CAA32151.1; PID:g35844
A;Note: part of this sequence, including the carboxyl end of the mature protein, was con
R;Crevel-Thieffry, I.; Cotterill, S.; Schuller, E.
Biochim. Biophys. Acta 1122, 107-112, 1992
A;Title: Characterisation of a tryptic peptide from human placental ribonuclease inhibit
A;Reference number: S23933; MUID:93338217; PMID:1633192
A;Accession: S23933
A;Molecule type: protein
A;Residues: 174-195 <CRE>
R;Nadano, D.; Yasuda, T.; Takeshita, H.; Uchide, K.; Kishi, K.
Arch. Biochem. Biophys. 312, 421-428, 1994
A;Title: Purification and characterization of human brain ribonuclease inhibitor.
A;Reference number: S48636; MUID:94311593; PMID:8017455
A;Accession: S48636
A;Status: preliminary
A;Molecule type: protein
A;Residues: 2-14 <NAD>
R;Poustka, A.; Wellenreuther, R.; Mewes, H.W.; Weil, B.; Wiemann, S.
submitted to the Protein Sequence Database, March 2000
A;Reference number: 224374
A;Accession: T47188
A;Status: preliminary
A;Molecule type: mRNA
A;Residues: 1-16; <AAA>
A;Cross-references: EMBL:AL161967
A;Experimental source: adult testis; clone DKFZp434K249
C;Genetics: GDS:RNH
A;Gene: GDS:RNH
A;Cross-references: GDB:125274; CMIM:173320
A;Map position: 11p15.5-11p15.5
A;Note: DKFZp434K249.1
C;Superfamily: ribonuclease inhibitor; leucine-rich alpha-2-glycoprotein repeat homology
C;Keywords: blocked amino end; duplication
F;34-433/Region: leucine-rich 57-residue repeats
F;57-8;/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR1>
F;114-138/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR3>
F;174-195/Region: inhibitory
F;228-252/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;285-309/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR4>
F;399-423/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR5>

Query Match          6.6%;   Score 212;  DB 2;  Length 46;;
Best Local Similarity  29.2%;   Pred. No. 2.6e-08;
Matches    59;   Conservative   35;   Mismatches  104;   Indels   4;   Gaps   2;

Qy    404  CALRLISCRLVAAQEKKKKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHL 463
                |  |      ||  ||   |  ||||    ::::: |   ||   |    | |||  |
Db     3:  QVVRLDDCGL:---ZARCKIISSALRVNPALAELN:RSNE:GDVGVHCVLCGLQTPSCK: 87

Qy    464  SSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTV 523
                ||  | ||  ||   |:  | |||   |::  |  | | |  | |: |:  | |
Db     88  QKLSLQNCCLTGAGCGVLSS:LRTLPTLQELHLSDNLLGDAGLQLLCEGLLDPQCRLEKL 147

Qy    524  RVQLPD-PQRGLQYLVGMLFQSPALTTLDLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLA 582
                || |    |::  |  | :  |                       |   | | |
Db    148  QLEYCSLSAASCEPLASVLFAKPDFKELTVSNNDINEAGVRVLCQGLKDSPCQLEALKLE 207

Qy    583  SVELSEQSLQELQAVKRAKPD: 604
                :  |    |  |
Db    208  SCGVTSDNCRDLCGIVASKaSL 229

RESULT 7
S20597
ribonuclease inhibitor - rat
C;Species: Rattus norvegicus (Norway rat)
C;Date: 22-Nov-1993 #sequence_revision 10-Nov-1995 #text_change 21-Jan-2000
R;Kawanomoto, M.; Motojima, K.; Sasaki, M.; Hattori, H.; Goto, S.
Biochim. Biophys. Acta 1129, 335-338, 1992

A;Title: CDNA cloning and sequence of rat ribonuclease inhibitor, and tissue distribut
A;Reference number: S20597; MUID:92162755; PMID:1536887
A;Accession: S20597
A;Molecule type: mRNA
A;Residues: 1-456 <KAW>
A;Cross-references: EMBL:X62528; NID:g57670; PIDN:CAA44368.1; PID:g5767:
C;Superfamily: ribonuclease inhibitor; leucine-rich alpha-2-glycoprotein repeat homolc
F;280-364/Domain: leucine-rich alpha-2-glycoprotein repeat homology <LRR>

Query Match          6.5%;   Score 209.5;  DB 2;  Length 456;
Best Local Similarity  24.3%;   Pred. No. 3.9e-08;
Matches    93;   Conservative   45;   Mismatches  129;   Indels  115;   Gaps  15;

Qy    259  GVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRWVQGGGQ 318
                |||  |        |||:    |  ||        :   :
Db     67  GVGLVLQGLQNPTCKI----------QKLSLQ--------NCSLTE----------A 95

Qy    319  GCPGVAPEVTEGAKGLED---TEEPEEEGEEPNYPLELLYCLYETQEDAFVRQALC-- 373
                   |  |                :    | | |:                :
Db     96  GC-GVLPDVLRSLLSTLRELHLNDNPLGDE-------GLKL---------LCEG 131

Qy    374  -RFPELALQRVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQASL 432
                 |   |       ||  |      |: |  |  |   :  || ::  :
Db    132  LRDPQ------CRLEKIQLEYC------NLTATSCEPLASVLRVKPDF-KE:VLSN 174

Qy    433  GGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALT 492
                     |     |: :      || |  | ||  |    | |
Db    175  NDFHEAG------------IHTLCQGLKDSACQLESLKLENGITSANCKDJCDVVASKASLQ 225

Qy    493  ELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVQLPD-PQRGLQYLVGMLRQSPALTT:D 551
                |  | |  | |:| |  | | |  |  |        :     |  |  |
Db    226  ELDLGSNKLGNTGIAALCSGLLPSCRLRTLWLWDCDVTAEGCKDLCVLRAKQSLKELS 285

Qy    552  LSGCQLPAPMVTYLCAVLQHQGCGLQTL--------------SLASVELSE 588
                |:|                                        : :::
Db    286  LAGNELKDEGAQLLCESLLEPGCQLESLWVKTCSLTAASCPHFCSVLTQNSSLFELQMSS 345

Qy    589  QSLQELQAVKRAK---PDLV: 606
                  :            |  |
Db    346  NPLGSGVVELCKALGYPDTVL 367

RESULT 8
T31665
hypothetical protein COS1.4 - sea squirt (Ciona intestinalis)
C;Species: Ciona intestinalis
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C;Accession: T31665
R;Bird, A.P.; Clark, V.; Jones, S.J.; Leitgeb, S.; Lennard, N.; Tweedie, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z21050
A;Accession: T31665
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1:1004 <BIR>
A;Cross-references: EMBL:Z80904; PIDN:CAB02586.1
C;Genetics:
A;Introns: 30/1; 539/3; 593/3; 662/3; 728/3; 769/3; 817/2; 881/2; 960/1

Query Match          5.3%;   Score 170;  DB 2;  Length 1004;
Best Local Similarity  19.5%;   Pred. No. 0.00011;
Matches   142;   Conservative  105;   Mismatches  273;   Indels  208;   Gaps  28;

Qy      5  PQRLLFILDGADE--LPALGGPEAAPCTDPFEAASGARVLGGLLSKALLPTALLLVTTRA 62
                 |  |         |   |  |          |    |  :  | | | ::   : |
Db    145  PNVAIF-FDGLDEASLKELVGGYSICKLD--EKSKPVDIMKONLFNLALLPKAKIVVTSTP 201

Qy     63  AAPGRLQGRLCSFPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCFVP 122
                 | | |  | |        | :   |:::  |  | | |  |    |::  |: |
Db    202  DEMFNLQHCYRPTSIFEVLGFLEEAKNNLGTQLCGEKYPAIK--KILDQQPNLAHLCYLP 259

Qy    123  --FVCWIVCTVLRQQLELEGRDLSRTSKTTTSVVLLFITSVLSSAPV-ADGP--RLQGDLR 177
                   |  |  |     |      |  | | | ||    :|  |    ||        |: :|

```
Db  138 LGRDLSRTSKTTTSVYLL--FITSVLSSAPVADGPRLQGDLRNICRLAREGV---LGRR 191
        |||:  |                      ::: |||     | |
Qy  120 GGRELTWLKQEVPEVTLSPDYRTAMDKADELYD--RVDGLLHQLTLQSNQRIQALELVQT 177

Qy  192 AQFAEKELEQLELRGSKVQTLFLSKKELPG--VLETEVTYQFIDQSFQEFLAALSYLLE 248
                                        ||
Db  178 LEARESGLHQIEVWLQQVGWPALEEAGEPSLDMLLQAQGSFQELYQVAQEQVRQGEKFLQ 237

Qy  249 DGGVPRTA--------AGGVGILLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM 300
                                                            ||
Db  238 ----PLTGWEAAELDPPGARFLARLRAQLTEFSR-ALAQR------CQRLADAERLF-- 282

Qy  301 VSERVKQEALRWVQGGQGCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY 360
               | ||
Db  283 ---QLFREALTWAE--------EGQRVLAELEQ-----ERPGVVLQQLQ-LH 317

Qy  361 ETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSYCVRCCPAGQAIRLISCRLVAAQ-EK 419
        :                     ||:
Db  318 WT------------RHPD--LPPAHFRKM------WALATGLGSEAIRQ-ECRWAWARCQD 357

Qy  420 KKKSLGKRLQASLG----GGSSQGTKQLPASLLHP------LFQAMTDPLCHL- 463
        ||||     |
Db  358 TWLALDQKLEASLKLPVGSTASLCVSQVPAAPAHPPLRKAYSFDRNLGQSLSEPACHCH 417

Qy  464 SSLTLSHCKLPDAVCRDLSEALRAAP----AJTELGLLHNRLSEAGLRMLSEGLAW 515
                                          ||
Db  418 HAATIAACRRPEAGGALPQASPTVPPPGSSDPRSLNRLQL---VLAEMVAT 466

Qy  516 PQCRVQTVRVQL-------PDPQRGLQ---YLVGMLRQ---SPALTTLDLSGGQLP 558
                                     ||
Db  467 EREYVRALEYTMENYFPELDRDVPQGLRGQRAHLFGNLEKLRDFHCHFFLRELEACTRH 526

Qy  559 APMVTYLCAVLQHQ 572
            ||
Db  527 PPRVAY--AFLRHR 538
```

RESULT 10
S75472
GTP-binding protein hflX - Synechocystis sp. (strain PCC 6803)
N;Alternate names: protein slr1521
C;Species: Synechocystis sp.
A;Variety: PCC 6803
C;Date: 25-Apr-1997 #sequence_revision 25-Apr-1997 #text_change 02-Feb-2001
C;Accession: S75472
R;Kaneko, T.; Sato, S.; Kotani, H.; Tanaka, A.; Asamizu, E.; Nakamura, Y.; Miyajima, N
o, K.; Okumura, S.; Shimpo, S.; Takeuchi, C.; Wada, T.; Watanabe, A.; Yamada, M.; Yasu
DNA Res. 3, 109-136, 1996
A;Title: Sequence analysis of the genome of the unicellular cyanobacterium Synechocyst
is.
A;Reference number: S74322; MUID:97061201; PMID:8905231
A;Accession: S75472
A;Status: nucleic acid sequence not shown; translation not shown
A;Molecule type: DNA
A;Residues: 1-534 <KAN>
A;Cross-references: EMBL:D90911; GB:AB001339; NID:g1653083; PIDN:BAA18033.1; PID:d10:8
A;Note: the nucleotide sequence was submitted to the EMBL Data Library, June 1996
C;Genetics:
C;Gene: hflX
A;Start codon: GTG
C;Superfamily: translation elongation factor Tu homology
C;Keywords: GTP binding; nucleotide binding; P-loop
F;362-489/Domain: translation elongation factor Tu homology <ETU>
F;368-375/Region: nucleotide-binding motif A (P-loop)
F;486-489/Region: GTP-binding NKXD motif

Query Match    3.6%;  Score 116; DB 2; Length 534;
Best Local Similarity  23.6%;  Pred. No. 0.63;
Matches 110; Conservative 60; Mismatches 170; Indels 126; Gaps 25;

```
Qy  182 LAREGVL---GRRAQFABKEL-EQLELRGSKVQTLFLSKKELPGVLETEVTYQFIDQS 235
        | |   |   ||                  ||
Db  140 LAEEAILDLVESLEAGF-EREFIARQVEAQGERVLVMGLKTSELD------DQS 186
```

```
Db  260 INFLIVFCLLSNE----GSDI--KTWTQVLIFSMTRFVESSHLKGEVPLDKVGAEVV 311
                                   :|
Qy  178 NICRLAREGVIGRRAQFAEKELEQLELRGSKVQTLFLSKKELPG-----VLETEVTYQF- 232
                                                                |||
Db  312 KIACLAYKGLQQRKLVFEKTEFDDVKLADEMVTNFFHTYVDISSGIRIKILEGNKRSYFT 371

Qy  233 DQSFQEFLAA------------LSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVJ- 276
                                 :|
Db  372 HHIWQEFYAAVYLMLFVSYREFEQLKTIFED----TQWSVVKFMFGICNPHAYKQCKI 426

Qy  277 ------TTRFLFGL---LSAERMRDIERHFGCMVSERVKQEALRWVQ----GQ 316
              |||   :|
Db  427 IFPATMIKDYEKKKFLFSMVMESLSSAKSEDLIRRFGWLHEYNDDETSKKFKICCLRVGL 486

Qy  317 GQCPGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYCLY----------- 361
        | |
Db  487 KNGVPKGHLPEVKDLVYAIKSFTKPHKLRLRSNWTTTEVLETLLRGIHGTTT--TRFVIN 546

Qy  362 --CEDAFVRQALCRFPEIALQRVRFCRMDVAVLSYCVRCCPAGQAIRLISCRLVAAQEK 419
               :|
Db  547 NIEMKDSLMELLLLHLD-ANEKLRF--DDVTNLS----------LFQAMTDP 577

Qy  420 KKKSLGKRLQASLGGSGGSSQGTKQLPASLLHPLFQAMTDP------ 471
        ||||
Db  578 -----SYMESLSNAINQRSNKIQLDLWIH--QQLNDDDVVYLAGCLGN-SRLNMSHT 627

Qy  472 KLPDAVCRDLSEALRAAPAL----------TELGLLHNRLSEAG 505
            ||
Db  628 YISSDQCRVLKQAIEQPSIQVHQLYPDILSYLNVNRPIIRFDFNTSVYFVHDQQFESS 697

Qy  506 LRVLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGMLRQSPAL-TTLDLSGGQLPAP----- 560
                                     ||
Db  698 --------SKCWIGS------------RG------GKLEVGGCELVVPPGALE 714

Qy  561 -MVTYLCAVLQHQGCG:QT---LSLASVELSEQSLQELQ-----AVKRAKPDLVI 606
            ||
Db  715 KDVEIKLTASLSLESEFLETPTLQCELASLITLKKQVTIKLQTHVVLDKETIRRCKVTIVY 774

Qy  607 THPALDGH 614
            ||
Db  775 TRVTTTVH 782
```

RESULT 9
T17233
hypothetical protein DKFZp434I1216.1 - human (fragment)
C;Species: Homo sapiens (man)
C;Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C;Accession: T17233
R;Koehrer, K.; Beyer, A.; Mewes, H.W.; Gassenhuber, J.; Wiemann, S.
submitted to the Protein Sequence Database, September 1999
A;Reference number: Z18722
A;Accession: T17233
A;Status: preliminary
A;Molecule type: mRNA
A;Cross-references: EMBL:AL117435
A;Experimental source: adult testis; clone DKFZp434I1216
C;Genetics:
C;Note: DKFZp434I1216.1

Query Match    4.0%;  Score 130; DB 2; Length 914;
Best Local Similarity  21.7%;  Pred. No. 0.11;
Matches 133; Conservative 95; Mismatches 222; Indels 164; Gaps 33;

```
Qy   34 EAASGARVLGGLLSKALL---PTALLVTTRAAAP------GRLQGRL--CSPQCA 78
        |  ||                                       |:
Db   34 EVPSGLQ-LEQLPSQSLLTHIPTAGLPTSLGGGLPYCHQAWLDFRRRLEALLQNCQAACA 72

Qy   79 EVRGFSDKDKDKKYFYKFFDERRAERARVRFVKENETLF-ALCCVPFVCWIVCTVLRQQLE 137
             |                                       |
Db   73 LLCQGAIES------VKAVPQFMEPGEVGQLIQQTEVLMQQVLDSPWLAWLQC-------Q 119
```

```
Qy   236  FQEFLAALSYLLEDGGV------------------PRTAA--GGVGTLLRGDAQPHSHLVLTTR 279
                                               :::       ::::::::::::::::::::::
Db   187  FAEGLTELERLVDSAGAQVLQVMQQNRSKPHPQTVVGEGKVEELALAVQTTGSNLVVFDR 246

Qy   280  FLFGLLSAERMRDIERHFGCMVSERVK---------------QEALRWVQG 315
          :::::::::::::::::::::::::::                ::::::::
Db   247  D---LSAAQVRNLEQRCGVRVIDRTELILDIFAQRAQSRAGKLQVELAQLEYLLPKLVG 302

Qy   316  QGGQCPGVAPEVTEGAKGLEDTEEPEEEGEEPNYPLELLYCLYETQEDAFVRQALCRF 375
          :::::::::::::::                            :::::::::::::::
Db   303  RGQQMSRLGGGI-GTRGPGETKLETERRTIQSRIAKLQ--------KQVNEQSHR 349

Qy   376  PELALQRVRFCRMDVAVLSYCVRCCPAGQAL---RLISCRLVAAQEKKKKKSLGKR-QASL 432
          :::::::::::::::::::::::::::::::::   :::::
Db   350  SRLRNORQQQAVPTVAIVGY---TNAGKSTLNALTQADIYAADQ--------LFATL 396

Qy   433  GGGSSQGTTKQLPASLLHPLFQAM-DPLCHLSSLTLSHCKLPDAVCRDLSEALRAA-PAL 491
          ::::::::::::::::::::::::   ::::::::::::::::::::::::::::::::
Db   397  -----DPTTRRJ-SLLDPENQTY-HPILLTDTVGFIH-KLPDA---LVDAFRATLEEV 443

Qy   492  TELGLLHNRLSEAGLRMLS-EGLAWPQ-CRVQTVRVQLPDPQRGLQYLVGMLRQSPALT 548
          ::::::::::::::::::  ::::::   ::::::
Db   444  TEADLL----LQVVDLSDRAWRRQIASVANILAEMPLATAPMVMVFNKIIDQVPSEA 495

Qy   549  TLDLSGCCLPAPMVTYLCCAVLQHCGCGLQTLSLASVE-SEQSLQEL 594
          ::::::::::::::::::::::::::::::::::::::   :::::::
Db   496  ---LAAAQADYPEAIFLAAA---QGIGLETL---KKR-LEQILTDL 532


RESULT 11
T35861
probable large secreted protein - Streptomyces coelicolor
C;Species: Streptomyces coelicolor
C;Date: 05-Nov-1999 #sequence_revision 05-Nov-1999 #text_change 05-Nov-1999
C;Accession: T35861
R;Saunders, D.C.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.
submitted to the EMBL Data Library, April 1999
A;Reference number: 221591
A;Accession: T35861
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-877 <SAU>
A;Cross-references: EMBL:AL049727; PIDN:CAB49562.1; GSEDB:GN00070; GSEDB:SC9B1.16c
A;Experimental source: strain A3(2)
C;Genetics:
A;Gene: SCOEDB:SC9B1.16c

Query Match      3.6%;   Score 114.5;  DB 2;  Length 877;
Best Local Similarity   22.0%;  Pred. No. 1.5;
Matches  113;  Conservative  63;  Mismatches 174;  Indels 163;  Gaps  27;

Qy   114  TLFALCFVPFV----CWIVCTV-------LRQQLELGRDLSRTSKTTSVYLLFITSVL 163
          ::::::::::    ::::::::        ::::::::::::::::::::::::::::
Db    21  SLILLALVPGVTLAAVWGV-TIQMFSEGLRLRAQTQLSRDTG--AXVGTEATLALQKERSI 78

Qy   162  SSAPVA--DGPRLQGDLRNLCRLAREGVLGRRAQFAEKEL----EQLELRGSKVQTLFLS 215
          :::::    ::::::::::::::::::::            ::::
Db    79  SAAWLAAPDGDRSELDAQR--------KKTDAAVAQLVGQSDAIESAPTRVSDRLYS 227

Qy   216  KKELPGVLE-------TEVT-YQFIDQ--------SFQEFLAALSYLLEDGGVPRT 255
          :::::::::        ::  :::::         :::::::::::::::::::::
Db   128  VLGSVGSLEYYRNQVDDPSDITAEQALDQYTSIVDEQIHAFQE----LSQ-VDDGDLTSQ 182

Qy   256  AAGGVGTLLRGDAQPHSHLVLTTRFLFGLLSAERMRDIERHFGCMVSERVKQEALRW-VQ 314
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::  ::
Db   183  AGPLVALEHAAELVSQEDARLLAWPSGRMDEEQW---SRFAQLVHTR---RWLVQ 232

Qy   315  GQGQCPGVAPEVTEGA-------KGLE-TE---EPEEEGEEPNYP----- 352
          ::::::::::::::::        :::  ::   :::::::::::
Db   233  DQ----TVPSLTGSAKTQTERILASSEWKSLQDVEDQVLEARSAGRGDRIDLPDARQR 286

Qy   353  ----LELLYCLYETQEDAFVRQALCRFPELALQRVRVRFCRMDVDAVLSYCVRCCPAGQALRL 408
```

```
Db   287  WITAFEKI---STQYGQLIRQQTDGLLDRSAEEARGLLIKAGVLS------AGGLIAL 335
          ::::::::   :::::::::::::::::::::::::::::::::::       :::::::
Db   409  ISCRLVAAQEKKKKSLGKRLQASLGGSSQGTTKQLPASLLHFLFQAMTDP-CHLSSLTL 468
          ::::::::::::::::::::::::::::::::::::::::::::::::::  ::::::::::
Db   336  LLC--IVMSWRITRSLSRRLRG---------------------------RLATLSL 363

Qy   469  SHCKLPDAVCR----DLSEALRAAPAL---TELGLLHNRLSEAGLRMLSEGLAWPQCRV 520
          :::::::::::    :::::::::::::   ::::::::::::::::::::::::::::::
Db   364  AEERLPDVVARLERGETVDVESATPPLYGRDELGVAQAFNAA---------QRTAV 412

Qy   521  QTVRVQLPDPQRGLQ-YLVGMLRQSPALTTLDL 552
          :::::::::::::::  :::::::::::::::::
Db   413  HTA-VELADTRRGFQKVIGIARQSQNLVNMQL 444


RESULT 12
T31668
hypothetical protein COS1.5 - sea squirt (Ciona intestinalis)
C;Species: Ciona intestinalis
C;Date: 29-Oct-1999 #sequence_revision 29-Oct-1999 #text_change 20-Jun-2000
C;Accession: T31668
R;Bird, A.P.; Clark, V.; Jones, S.J.; Leitgeb, S.; Lennard, N.; Tweedie, S.
submitted to the EMBL Data Library, October 1996
A;Reference number: Z21050
A;Accession: T31668
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1075 <BIR>
A;Cross-references: EMBL:Z80904; PIDN:CAB02589.1
C;Genetics:
A;Introns: 48/1; 557/3; 611/3; 685/2; 721/2; 739/1; 779/3; 820/3; 853/2; 914/2; 993/1

Query Match      3.5%;   Score 114;  DB 2;  Length 1075;
Best Local Similarity   19.0%;  Pred. No. 2.2;
Matches  126;  Conservative 102;  Mismatches 278;  Indels 158;  Gaps  27;

Qy    41  VLGGLLSKALLPTALLVTTRAAAPGRLQGRLCSPQCAEVRGFSDKDKKKYFYKFFRDER 100
          :::::::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   198  IMKNLFNLTLLPKAKIVVTSTLHQMYKLHPDYRPTSIFEVLGLLEEAKNNLGTQLCGEKY 257

Qy   101  RAERAYRFVKENETLFACCFVP--FVCWIVCTVLRQQELGRDLSRTSKTTSVYLLFIT 158
          ::::::::::::::::::::::   ::::::::::::::::::::::::::::::::::
Db   258  PAIK--KILDQQPNLAHLCYLPINFILIVFCLLSNE---GSDI---KTMIQVLIFSMT 307

Qy   159  SVL------SSAPVADGPRLQGDLRNLCRLAREGVLGRRAQFAEKELEQLELRGSKVQTL 212
          :::       ::::::::::::::::::::::::::::::::::::::::::::::::::
Db   308  RFVELSHLKGEVPL--DRVGAEMVKLARLAYKGLCQRKLVFEKTDFDDVKLADEMVTNF 364

Qy   213  FLSKKELPG------VLETEVTYQFIDQSFQEFLAA------LSY-------LLEDGGV? 253
          :::::::::       :::::::::::::::::::::      :::        :::::::
Db   365  FHTVVDKSSGIRMKILEGNKRSYFTHLIWQEFYAAVYLMLFVSYREFEQLKPIFKDAQWK 424

Qy   254  RTAAGGVGTLLRGDAQPHSHLVLTTRF-----------------LFGLLSAERMRDIER 295
          ::::::::::::::::::::::::::::                 ::::::::::::::::
Db   425  RV----VGFMF-GICNPPAYKQLKJVFPATMIKDYEKKELMVPMMESLWSARGEDLIR 479

Qy   296  HFGCMVSERVKQEALRWVQGQGQCPGVAPEVTEGAKGLED-------TEEPEEEGEE 349
          :::::::::::::::::::::::::::::::::::::::::        ::::::::::
Db   480  RFG-------WLHEYND--------DESSKKFPEDYLPVGLKMDAPKHLSEVKD 517

Qy   349  PNY-------PLELLYCLYETQEDAFVRQALCRFPELALQRVRF----CRMDVAVLSYCV 397
          :::        :::::::::::::::::::::::::::::::::    :::::::::::::
Db   518  LVYALKSFTKPHKLRLDSYETTTTEVLETLLRGVHGTTTTITRFVINNIEMKDSLMELLL 577

Qy   398  RCCPAGQALRLISCRLVAAQEKKKKKSLGKRLQASLGGSSQGTTKQLPASLLHPL----- 452
          ::::::::::::::::::::::::::::::::::::::::::::::::::::::::
Db   578  LHLDAMEELRFIRVTNLSPYMER-------------LSNAINQRSNKIQLVLIKHKLHDDEYD 626

Qy   453  FQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPAL-----------TELGLLHNRL 501
          ::::::::::::::::::::::::::::::::::::::::            ::::::::::
Db   627  VKYLAGCLGNISLLYMWGTDISSDQCSVLKQATCQLPSIQVHQLYPDILSTYLNVARNWI 686
```

```
Qy   502  SEAGLRMLSEGLAWPQCRVQCTVRVQLPDPQRGLQYLVGMLRQSPALTTDLSGCQLPAP-  560
                |:|  ||| || ||:|:|  | |  |||:||:| :||  ::|
Db   687  SN-------------WFVSEVDLVPEAL-NLGNGLKDSKGQPRRREELCSQDVS---VPAPT  73:

Qy   561  NVTY--LCAVLQHGGCGLCTLSLA----SVE-SEQSLQELQAVKRAKPLVITHPAL  611
              ||:  |  | ||:| |: :|        :|  ||:: :|:|:|: ||| :|:|
Db   561  HLRRLNPEGKPLPILRIGEGIENLVRESIRQTAMGTYALSSRHKQILQLIEQALKQS  623

Qy   732  QVNYNHYCNAPYHQ---VRFLTQAFILFMTNSLNLNHPNGLDQVEKIWRLEDSEYLETPKJ  788
                              |:|  || ||: ::||:|   | ||  ||:||:| |||:||
Db   545  PAL---TTLDLSGCQLPAPMVT-YLCAVLQHGCGLQTL--SLASVELSEQSLQE  593

Qy   612  DGHP  615
           ||||
Db   624  AKLFIVTSVDTRRFLRKITEATLFDVPILSWQLGEESLIQVESIDSEEELAD  678

Qy   789  QCEP  792
           ||||
Db   516  IAETLQRLVSERVSIRDLRLIFGTLIDWAPREKDVLMLTEYRIA------------LRR  563
```

RESULT 13
AF0697
probable type III secretion protein ssaV [imported] - Salmonella enterica subsp. enteric
C;Species: Salmonella enterica subsp. enterica serovar typhi
C;Note: this species has also been called Salmonella typhi
C;Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 18-Nov-2002
C;Accession: AF0697
R;Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Churcher,
S.; Moule, S.; O'Gaora, P.
C;Nature 413, 848-852, 2001
A;Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.;
A;Title: Complete genome sequence of a multiple drug resistant Salmonella enterica serov
A;Reference number: AB0502; MUID:21534947; PMID:11677608
A;Accession: AF0697
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-681 <PAR>
A;Cross-references: GB:AL513382; PIDN:CAD01951.1; PID:gi6502793; GSPDB:GN00175
C;Genetics:
A;Gene: ssaV
C;Superfamily: regulatory protein icrD

Query Match       3.5%;    Score 113.5;   DB 2;   Length 681;
Best Local Similarity  21.8%;    Pred. No. 1.3;
Matches  143;    Conservative   86;   Mismatches   243;   Indels  183;   Gaps   33;

```
Qy    36  ASGARVLGGLSKALLPTALL------LVTTRA----AAPGR---LQGRLC  73
           |  :  |  |:| ||  ||       |||:       ||:|    | |
Db   141  DLSRTSKTTSVYLLFITSVLLSSAPVADGPRLQGDLRNLCRLAREGV.GRRAQFAEKELE  200

Qy   110  AFGKFVVGGNLTVGLVVFTITIVQFIVITKGIERVAEVSARFSLDGMPGKQMSIDGDLR  169
           || ::| |:|  : | | |||| |||| |||  |||  |:| ||  |    |: ||||
Db   284  DIEQASQAVTT--------LSKTPVAAVEMDG--RALRSVADK--KGMPEFIAKLDLE  330

Qy    74  SPQCAEVRGFSDKDKKKYFYKFFRDERR---AERAYRFVKENETL--FALCSVPFVCWI  127
              |  |  ||   |:|| : |||||     |: |  :|:||     |   ||: : |
Db   201  Q-----LELRGSKVQTLFLSKKELPGVLETEVTYQFIDQ-----RYHIIESVPF*SESKTVATLWGIRKGMFP  387

Qy   170  A-------GVIDADHARTLRQHVQQESRFLGAXDGAXKFVK-GDTIAGIVV-VNIIGGI  221
                    |   |||  |||:  |   :|||  | ||   |  ||:|  ||  ||| :||
Db   331  AAALLIESHASDAQTLHAQCEQVMSALQ--RYHIISVPF*SESKTVATLWGIRKGMFP  387

Qy   128  VCTVLRQQLELGRDLSRTSKTTTSVYLLFITSVLSSAPVADGPRLQGDLRNLCRLAREGV  187
           | |:|||  ||  | ||:| ||:||||||||||||||||||||||||||||||||||||
Db   244  -----SYLLEDGVP-RTAAGGV----------------GTLLRGDAQPSHL  274

Qy   222  IIAIVQY------DMSMSEAVHT------YSVLS----SGDG-LCGQIFSLLISLSAGI  263
           ::|||          |: |||        ||||       | |  | :  |:|:|:|
Db   388  AVGAVREVGTTVIIEDVAFPVEKLAAGVRDLQALFDKYHYNEAIIFGHALEGNL-----HF  443

Qy   188  L------GRRAQFAEKELEQLELRGSKVQTLFLSKKE----FGVLETEVTYQFIQSF  236
           |       |: |:  ||:|| || ::|  ||||| | | ||| ||:| ||||| |
Db   275  VLTTRFLGLLSAERMRDIERHFGCM--VSERVK--QEALRWVQGQGQCPGVAPEVTE  329

Qy   264  IVTRVFGEKRQNLATELSSQIAR--QPQSLILTAVVLMLLALIPG---------FPFITLAF  314
           |:| |||:  | | || |  :|   || ||| | ||  | | ||          || |
Db   444  VFTQGF----DKQSIERYGAFMDDVAELVAVKYQGSLKAEHGTGR---NMAPYV-  491

Qy   237  QEF.AALSYLLEDGVPR.AAGGVGTLLRGDAQPSHLVLTTRFLFGLLSAERMRDIERH  296
           ||: | |:  | | |         |  |  ||  | |:| ||||| |||:|:||||
Db   330  GAKGLEDTEEPEEEGEEPNYPLELLYCLYETQEDAFVRQALCRPPELALQRVRFCR-  387

Qy   315  FSALLALPIILIRRKKSVVSANGIEAPEKDSMVPGA-CFLILELTPTLHSADLIRDI---  370
           |  | |:  |||:||   |  | : |      |||  | |:|||  | || || |:
Db   492  -------ELEWGKEG--------YALMQKIKALFDPNRLLN-PGVIINEDKRHSHES  531

Qy   297  FGCMVSERVKQEALRWVQGQGQCGCPGVAPEVTEGAKG.EDTEEPEEEE---GEEPNVPL  353
           |   || ||    |:    |   ||  | || || :    |  :|||       ||||
Db   388  ----MDVA--VLSYCVRC-----CPAGQALRLISCRVAAQEKKKSLGKRLQASLGGG  435

Qy   371  -------DAMRWFLFEDTGVP--LPEV-------NIEVLPEPTEKLTVLLYOEPVTSL  412
                   | | | |:|| ||  ||:        ||:|||| ||||| ||||||:||
Db   532  NLKPMPAADNLVDRCIECGFCEPVCPS------------RTLTLSPRQRIVLYRELQRRRAAG  582

Qy   354  ELLYCLYETQEDAFVRQA---------LCRPSPELALQRVRFCRMDV------  390
           || |:||||||||| |||          |: |  |:||:|| | |:
Db   436  SSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHC-------KLPDA----V  477

Qy   413  SI-----PAQADYLLIGADASVGDSQT-PNGMGGICWUSKDMAHKAQGFGLDVFAGSQR  467
           |:      ||  | | |  | |  |    |    |    |  | ||    |: : ||::
Db   583  ENVASSELEQVPEYQGLDTCAATGLCAERCPVGINTGDLVKLRIAKYQKFTPI  636

Qy   391  -AVLSYCVRCCPAGQALRLISCR-LVAAQEKKKKSLGKR-CASLGGGSQGTTKQLPASL  448
               || |:|| ||| ||  |:|  | |  |||| |    |   ||  ||  | |
Db   478  CRDLSEALRAAPALTELGLLHNRLSEA--------GLRMLSEGLA--W---PQCR  519

Qy   468  ISALLKCVLRHMGEFIGVQETRYLMNAVEKNYSELVKELQ--------RQLPINK  515
           ||| ||  | : | | |  ||:|::| |  |    | |||         |::| :
Db   637  ARWTAEHFSATTLARGGKANQLATQVLGEKSVDSMVNGLRRISKGKTPLAMPEMPQAN  696

Qy   449  LHPLFQAMTDPLCHLSSLTLSHCKLPDAVCRD----LSEALRAAPALTELGLLHNRLSE  503
           |||||||| ||| ||||||||:|| ||||  |     |||: | | |||:| ||| |:
Db   520  VQTVRV---QLPDPQRGLQYLV-----GMLRQSPA-------LTTLDLSGCQLPAPM  561

Db   697  THSLEAVENLPRSDKKVVYLPSCASRNMGQQASATDQRPLTEVTLSLLNKAGFEVILP-  755
```

RESULT 14
C82392
oxidoreductase/iron-sulfur cluster-binding protein VCA0985 [imported] - Vibrio cholera
C;Species: Vibrio cholerae
C;Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C;Accession: C82392
R;Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J
...chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers,
..R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000
A;Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.
A;Reference number: A82035; MUID:20406833; PMID:10952301
A;Accession: C82392
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-959 <HEI>
A;Cross-references: GB:AE004425; GB:AE003853; NID:g9658424; PIDN:AAF96881.1; GSPDB:GNC
A;Experimental source: serogroup O1; strain N16961; biotype E* Tor
C;Genetics:
A;Gene: VCA0985
A;Map position: 2

Query Match       3.5%;    Score 112.5;   DB 2;   Length 959;
Best Local Similarity  20.8%;    Pred. No. 2.4;
Matches  129;    Conservative   83;   Mismatches  200;   Indels  207;   Gaps   35;

```
Qy   141  DLSRTSKTTSVYLLFITSVLLSSAPVADGPRLQGDLRNLCRLAREGV.GRRAQFAEKELE  200
           | :    ::||  |:  ||| |:||| | || |:||||:|||||||   | :||:|||:|
Db   284  DIEQASQAVTT--------LSKTPVAAVEMDG--RALRSVADK--KGMPEFIAKLDLE  330

Qy   201  Q-----LELRGSKVQTLFLSKKELPGVLETEVTYQFIDQ--RYHIIESVPF*SESKTVATLWGIRKGMFP  387
               || |:|| || ::|  ||||| |||:| |:|::|     | ::|: :|
Db   331  AAALLIESHASDAQTLHAQCEQVMSALQ--RYHIISVPF*SESKTVATLWGIRKGMFP  387

Qy   244  -----SYLLEDGVP-RTAAGGV----------------GTLLRGDAQPSHL  274
                ||| | | |   |||| |                |  |  ||  |
Db   388  AVGAVREVGTTVIIEDVAFPVEKLAAGVRDLQALFDKYHYNEAIIFGHALEGNL-----HF  443

Qy   275  VLTTRFLGLLSAERMRDIERHFGCM--VSERVK--QEALRWVQGQGQCPGVAPEVTE  329
           | :  ||:||||| |:||||||  |   || ||   ||| |  | |  | |  |||
Db   444  VFTQGF----DKQSIERYGAFMDDVAELVAVKYQGSLKAEHGTGR---NMAPYV-  491

Qy   330  GAKGLEDTEEPEEEGEEPNYPLELLYCLYETQEDAFVRQALCRPPELALQRVRFCR-  387
           | | ||:| || |         ||  |:|||:||||:|||| ||||||||:|| ||
Db   492  -------ELEWGKEG--------YALMQKIKALFDPNRLLN-PGVIINEDKRHSHES  531

Qy   388  -MDVA--VLSYCVRC-----CPAGQALRLISCRVAAQEKKKSLGKRLQASLGGG  435
            | |   |||| |:||     ||  | || || ||  |||| |   |   ||
Db   532  NLKPMPAADNLVDRCIECGFCEPVCPS------------RTLTLSPRQRIVLYRELQRRRAAG  582

Qy   436  SSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHC-------KLPDA----V  477
           ::  | | |  |:|||||||| ||| ||||||||:|        ||||    |
Db   583  ENVASSELEQVPEYQGLDTCAATGLCAERCPVGINTGDLVKLRIAKYQKFTPI  636

Qy   478  CRDLSEALRAAPALTELGLLHNRLSEA--------GLRMLSEGLA--W---PQCR  519
           |   |||: | | |||:| ||| |:         || ||:|| |     |::
Db   637  ARWTAEHFSATTLARGGKANQLATQVLGEKSVDSMVNGLRRISKGKTPLAMPEMPQAN  696

Qy   520  VQTVRV---QLPDPQRGLQYLV-----GMLRQSPA-------LTTLDLSGCQLPAPM  561
           |:: ||    |:  : | || |      | ||           || |||| | |
Db   697  THSLEAVENLPRSDKKVVYLPSCASRNMGQQASATDQRPLTEVTLSLLNKAGFEVILP-  755
```

```
Qy    562  VTYLCAVLQHGCGLQTLSLASVELSEQSLQEL------QAVKRAK-PDLV-THP----AL  615
                |:|:  |||   |   ||||:| : :||       |     ||  |  |::|||
Db    756  AELSSQCCGMPYDSK3MTEIAQSKAQQLEHALWQASCEGRYPILMDTSPCAKRSI  810

Qy    612  DGHPCP------PKELISTF  625
                  |:|  |||
Db    811  EQFTKPMEILEFTGLVSRY  829
```

RESULT 15
T03465
probable exonuclease (EC 3.1.15.-) - Rhodobacter capsulatus
C;Species: Rhodobacter capsulatus
C;Date: 24-Mar-1999 #sequence_revision 24-Mar-1999 #text_change 02-Jun-2000
C;Accession: T03465
R;Vlcek, C.; Paces, V.; Maltsev, N.; Paces, J.; Haselkorn, R.; Fonstein, M.
Proc. Natl. Acad. Sci. U.S.A. 94, 9384-9388, 1997
A;Title: Sequence of a 189-kb segment of the chromosome of Rhodobacter capsulatus SB1003
A;Reference number: Z14955; MUID:97404404; PMID:9256491
A;Accession: T03465
A;Status: preliminary; translated from GB/EMBL/DDBJ
A;Molecule type: DNA
A;Residues: 1-1238 <VLC>
A;Cross-references: EMBL:AF010496; NID:g3128256; PIDN:AAC16118.1; PID:g3128266
C;Genetics:
A;Gene: sbcC
A;Map position: 1
C;Superfamily: chromosome segregation protein SMC1
C;Keywords: DNA repair; exonuclease; hydrolase

```
Query Match            3.5%;   Score 112.5;   DB 2;   Length 1238;
Best Local Similarity 21.0%;   Pred. No. 3.4;
Matches  152;  Conservative  68;  Mismatches  316;  Indels  187;  Gaps  28;

Qy      4  QPQRLLFILDGACELPALGGP--------EAAPCTDPFEAASGARV------------LGGLLSK   48
           | |  |   |  |  |                                           ||||| |
Db    274  QVQAAAAALAGAEALQAEAGPARAQLDRLERAAPLHLPWQRAAEARARLDAARSGAEAAA  333

Qy     49  ALLPTALLLVTRAAAPGR.QGRLCSPQCAEVRGFSDKDKKYFYKFFRDERRAEEAYRF  105
            | :   |   |  |||    | |     :  :||   |   |   |:  | |
Db    334  ASRSLAAQALATRADAAAR-------ATAAGTETEEAFKAFGRIW-DRAAAALDAQIA  382

Qy    109  VKENETLFALCFVPFVCWIVCTVLRQQLELGRDLSRTSKTT7SVLLFI7SVLSSAPVAC  168
               |          |    | |       ||  |    || | :||  | ||
Db    383  TRATEAEAARTRAAETARAAAGIRRAEADLAAAETRACQARQAAEAR-LAE-AAQAPLAD  441

Qy    169  G-PRLQCDL--RNLCRL--------------------------------AREGVLGRR  191
           | |  | |   | | |                                  |   ||:
Db    442  DWFQLRRDLADHRAACAACAQAAAATAAQAAAQDRAQALAQEARAAAAEARDQAADSALAKA  501

Qy    192  AQFAEKE.EQLELRGSKVQ7LFLSKKELPGVLETEVTYQFIQSFQEFLAALSYLLEDGG  251
           |  | |   | |                          ||    |||  |  | || :
Db    502  AAALQQQAAPLEAAHPPARTADLAR----------IE7D---------LAELTRALRDGA  542

Qy    252  V----------PRTAAGGVGTLLRGDA--------QPHSHLVLIT7RFL----FGLLSAER  289
                    |  |  | : | |                   |  || : |       |  | ||
Db    543  EAAALGAAAQRAATAAAETAAALRAAQESAAKQDLDRAETOIAALTAPLEQADLALSDAAR  602

Qy    290  MRDIERHFG--CMVSERVKQ----------EALRMVCGQGCQCPGVAPEVTEGAKGLE  335
           |   |  | |                          |:  |   |  |  | | | | :
Db    603  SLRAQLSAGSPCPVCGALEHPTPAEAGLAHLAERLRADQAAARGAAQAARDALTAACGAR  662

Qy    336  DTEEPEEEGEEPNYP---LELLYCLY-ETQEDDAFVRQALCRFP------ELALQRVRF  385
                 || |      |      :||    | ||      |||
Db    663  ATAEAARGTQAAEDQRRAQTRAEEAAAAWGDTQPRVSARPLAPALPGTPDPTALAAAQDRL  722

Qy    386  CRMDVAVLSYCVRCCPAGQA-LRLISCRLVAAQEKKKKSLGKRLQASLGGGSSCGTTKQL  444
                    |   |   | | |   |   |   |  | |
Db    723  CAIQTAE------AAAQAEISALRTRLTEA-ERCRER:RRALLA-----HRGTRERL  767

Qy    445  PASLLHPLFQAWTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTELGLLHNRLSEA  504
```

```
Db    768  AVQ-------QAET----AQEAALAEARTEAAARRDGLALALAPLLARAG-----EDDP  811
                                |:| || |  |||       |
Qy    505  GLRMLSEGLAWPQCRVQTVRVQLPDPQRGLQYLVGML-----RQSPALTTLD  551
            | |  |  |     |  ||   |  |      |:  |       :|||| ||
Db    812  AAPGLAERLAATVSAVGAARTGLQAAQEALSALAPLQLAAARRDSETATAQCAQSAQAARD  871
                 |   |    |  |    |     |    :|:  :|
Qy    552  LSGCQLPAPMVTYLCAVLQHQGCCL---QTLSLASVELSEQSLQELQAVKRAKPDLVITH  608
             ||   |  |   |  |  ||:   :|| | |
Db    872  -WAALRAERAPLLDGQPTALHRSRFNDQRLAAQRAQQAAAADLATAQ  921

Qy    609  PAL  611
           :||
Db    922  AAL  924  .
```

Search completed: October  2, 2003, 17:52:44
Job time : 47 secs

CM nucleic - nucleic search, using sw model

```
             GenCore version 5.1.6
             Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:      October 3, 2003, 03:41:16 ; Search time 696 seconds
                                         (without alignments)
                                         10429.294 Million cell updates/sec

Title:          US-10-029-347-1
Perfect score:  2689
Sequence:       1 cggacgcgtgggcgcgcagc........aaaaaaaaaaaaaaaaaaaa 2689

Scoring table:  IDENTITY_NUC
                Gapop 10.0 , Gapext 1.0

Searched:       252756 seqs, 134919190:7 residues

Total number of hits satisfying chosen parameters:   5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing:  Minimum Match 0%
                  Maximum Match 100%
                  Listing first 45 summaries

Database :  N Geneseq 19Jun03:*
        1:  /SIDS17/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
        2:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
        3:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
        4:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
        5:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
        6:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
        7:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
        8:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
        9:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
       10:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
       11:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
       12:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
       13:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
       14:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
       15:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
       16:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
       17:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
       18:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
       19:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
       20:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
       21:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
       22:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
       23:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
       24:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
       25:  /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*
```

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB ID | Description |
|---|---|---|---|---|---|
| 1 | 2689 | 100.0 | 2689 | 24 ABS63485 | DNA encoding human |
| 2 | 2626 | 97.7 | 3365 | 22 AAD08836 | Human G-protein co |
| 3 | 2605 | 96.9 | 2763 | 24 ABS63486 | Human leucine-rich |
| 4 | 2020.0 | 75.1 | 2054 | 24 ABS63487 | Human leucine-rich |
| 5 | 1938.6 | 72.1 | 2109 | 24 ABX97042 | Human NOV18a cDNA. |
| 6 | 1571.2 | 58.4 | 5556 | 24 AAL47132 | Pyrin domain conta |
| 7 | 1571.2 | 58.4 | 6461 | 22 AAS00033 | Human ATLAS-2-enco |
| 8 | 1368.6 | 50.9 | 1795 | 22 ABA08979 | Human vasopressin |
| 9 | 832.8 | 31.0 | 2251 | 14 AAQ38723 | AII/AVPv2 receptor |
| 10 | 798.6 | 29.7 | 933 | 22 AAK83313 | Human immune/haema |
| 11 | 740.6 | 27.5 | 989 | 22 AAK62741 | Human immune/haema |
| 12 | 729.6 | 27.1 | 990 | 24 ABL90582 | Human polynucleoti |
| 13 | 464.6 | 17.3 | 704 | 24 ABQ29970 | Oligonucleotide fo |
| 14 | 464.6 | 17.3 | 704 | 24 ABQ29971 | Oligonucleotide fo |
| 15 | 432.6 | 16.1 | 704 | 24 ABQ29972 | Oligonucleotide fo |
| 16 | 432.6 | 16.1 | 704 | 24 ABQ29973 | Oligoruc-eotide fo |
| 17 | 427.6 | 15.9 | 678 | 22 ACA56413 | Norway rat signall |
| 18 | 289.2 | 10.8 | 463 | 22 AAK88542 | Human digestive sy |
| 19 | 277 | 10.3 | 3186 | 24 AAL44363 | Human PYRIN-8 cDNA |
| 20 | 273.4 | 10.2 | 3300 | 24 AAL47129 | Pyrin domain conta |
| 21 | 258 | 9.6 | 314 | 24 ABS63488 | Human leucine-rich |
| 22 | 236.6 | 8.8 | 257 | 25 ACA55610 | Norway rat signall |
| 23 | 226.4 | 8.4 | 3857 | 25 AAD14323 | Human PYRIN 1 (PYR |
| 24 | 226.4 | 8.4 | 3857 | 25 ABX93556 | Huma cDNA encoding |
| 25 | 195.4 | 7.3 | 2847 | 22 AAS01487 | Human secreted pro |
| 26 | 193.6 | 7.2 | 2847 | 25 ABZ73494 | Secreted protein-e |
| 27 | 193.6 | 7.2 | 1557 | 24 AAJ47143 | Pyrin domain-e |
| 28 | 179.2 | 6.7 | 3431 | 25 AAD14322 | Human nucleotide b |
| 29 | 179.2 | 6.7 | 3431 | 25 ABX93555 | Huma cDNA encoding |
| 30 | 179 | 6.7 | 4931 | 24 ABL59333 | Nucleotide sequenc |
| 31 | 177.6 | 6.6 | 3189 | 22 AAI67185 | Nucleotide sequenc |
| 32 | 177.6 | 6.6 | 3531 | 24 AAL47128 | Pyrin domain conta |
| 33 | 177.6 | 6.6 | 4787 | 23 AAS92566 | DNA encoding novel |
| 34 | 171.6 | 6.4 | 3085 | 23 AAS68111 | DNA encoding novel |
| 35 | 171.6 | 6.4 | 3150 | 22 AAI67184 | Nucleotide sequenc |
| 36 | 171.6 | 6.4 | 3263 | 24 AAI70684 | Human nucleotide b |
| 37 | 171.6 | 6.4 | 3459 | 24 ABQ78049 | Human CGDD encodin |
| 38 | 170 | 6.3 | 3069 | 24 AAI47134 | Pyrin domain conta |
| 39 | 167 | 6.2 | 3368 | 24 AAL44366 | Human PYRIN-3 cDNA |
| 40 | 165.4 | 6.2 | 2575 | 22 AAI70683 | Human nucleotide b |
| 41 | 165.4 | 6.2 | 2767 | 24 ABT16018 | NOVX related polyn |
| 42 | 165.4 | 6.2 | 3172 | 24 ABN99366 | Human secreted pro |
| 43 | 163.8 | 6.1 | 969 | 22 AAK78135 | Human immune/haema |
| 44 | 163.8 | 6.1 | 969 | 22 AAK78136 | Human immune/haema |
| 45 | 162.2 | 6.0 | 969 | 22 AAK78134 | Human immune/haema |

ALIGNMENTS

```
RESULT 1
ABS63485
ID   ABS63485 standard; cDNA; 2689 BP.
XX
AC   ABS63485;
XX
DT   15-NOV-2002 (first entry)
XX
DE   DNA encoding human leucine-rich repeat small intestine I (HLRRSI1).
XX
KW   Human; human leucine-rich repeat small intestine I; HLRRSI1; asthma;
KW   proliferative disorder; gastrointestinal disorder; renal disorder;
KW   neural disorder; reproductive disorder; calcium regulation; apoptosis;
KW   immune system; anaemia; human immune deficiency virus; HIV; cancer;
KW   blood coagulation disorder; autoimmune disorder; allergic reaction;
KW   inflammatory condition; cardiovascular disorder; ischaemia;
KW   neurological disorder; infectious disease; cytokine production;
KW   expressed sequence tag; EST; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200261086-A2.
XX
PD   08-AUG-2002.
XX
PF   20-DEC-2001; 2001WO-US049739.
XX
PR   22-DEC-2000; 2000US-257774P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
```

PI  Feder J, Ramanathan C, Mintier G;
XX
DR  WPI; 2002-619252/66.
DR  P-PSDB; ABG78454.
XX
PT  New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT  their fragments and homologues, useful for preventing, treating and
PT  ameliorating medical conditions, e.g. proliferative, gastrointestinal,
PT  or renal disorders -
XX
PN  Claim 1; Figure 1; 336pp; English.
PS
XX
CC  The invention relates to isolated nucleic acid molecules (I) encoding
CC  human leucine-rich repeat small intestine 1 (HLRRSI1) polypeptides.
CC  The nucleic acid molecules and polypeptides are useful for preventing,
CC  treating and ameliorating medical conditions, such as proliferative,
CC  gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC  related to aberrant calcium regulation or apoptosis modulation, either
CC  directly or indirectly. They are also useful for treating, preventing
CC  and/or diagnosing diseases, disorders and/or conditions of: immune system
CC  by activating or inhibiting the proliferation, differentiation, or
CC  mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC  anaemia; immunologic deficiency syndromes, e.g. human immune deficiency
CC  virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC  e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC  myasthenia gravis; asthma or allergic reactions; inflammatory
CC  conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC  e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC  ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC  Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC  pneumonia, or viral, bacterial, and fungal infections. The HLRRSI1
CC  polypeptides are useful for modulating cytokine production, antigen
CC  presentation, or other processes such as boosting immune responses.
CC  ABS63495-ABS63504 represent HLRRSI1 coding sequences and PCR primers of
CC  the invention.
XX
SC  Sequence 2689 BP; 534 A; 861 C; 844 G; 450 T; 0 other;

Query Match     100.0%;   Score 2689;  DB 24;  Length 2689;
Best Local Similarity   100.0%;  Pred. No. 0;
Matches 2689;   Conservative  0;  Mismatches  0;  Indels  0;  Gaps  0;

Qy      1  CGGACGCGTGGGCGCGGCGCCGCAGCCTGGCTGACCTGACCCAGTGCCCGGACCGGCGG   60
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  CGGACGCGTGGGCGCGGCGCCGCAGCCTGGCTGACCTGACCCAGTGCCCGGACCGGCGG   60

Qy     61  CGCCGGTGCGCCGCAGATGCTGCCCCAGCCGCGCTGCTTCATCCTGACCATCCTGACCGG  120
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  CGCCGGTGCGCCGCAGATGCTGCCCCAGCCGCGCTGCTTCATCCTGACCATCCTGACCGG  120

Qy    121  ACGAGCTGCCGGCGCGTGGGGGGCGCGGCGCCCGAGGCGGGGGG  180
           ||||||||||||||||||||||||||||||||||||||||||||
Db    121  ACGAGCTGCCGGCGCGTGGGGGGCGCGGCGCCCGAGGCGGGGGG  180

Qy    181  CGAGGCGGCGCGGGTGCGTAGGCGGGGCTGCTGAGTAAGGCGCTGCTGCCAGGCCCTCC  240
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  CGAGGCGGCGCGGGTGCGTAGGCGGGGCTGCTGAGTAAGGCGCTGCTGCCAGGCCCTCC  240

Qy    241  TGCTGGTGACCACGCGCGCGGCGGGG  300
           ||||||||||||||||||||||||||
Db    241  TGCTGGTGACCACGCGCGCGGCGGGG  300

Qy    301  AGTGCGCCGGAGTGCGCGGCGCGGGT  360
           ||||||||||||||||||||||||||
Db    301  AGTGCGCCGGAGTGCGCGGCGCGGGT  360

Qy    361  TCCGGGATGAGAGGAGGAGGAGGAGGCGCGGCCTACCCGTTGGTGAAGGAGAACGAGACGACGTGT  420
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  TCCGGGATGAGAGGAGGAGGAGGAGGCGCGGCCTACCCGTTGGTGAAGGAGAACGAGACGACGTGT  420

Qy    421  TCGGCGCTGTGCTTCGGTGCTTCGTGTGCTGGATCGGTGCCACCGTGCTGCGCGTGGCCCAGCAGC  480
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  TCGGCGCTGTGCTTCGGTGCTTCGTGTGCTGGATCGGTGCCACCGTGCTGCGCGTGGCCCAGCAGC  480

Qy    481  TGGAGCTCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCGCTTT  540
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGGCGCTCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCGCTTT  540

Qy    541  TCATCACCAGCGTCTTGAGCTCGGTCGGCTCCGGTAGCCGCGGGCCCGGTTGCAGGGGACC  600
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TCATCACCAGCGTTCTGAGCTCGGTCGGCTCCGGTAGCCGACGGGCCCGGTTGCAGGGGACC  600

Qy    601  TGCCAATCTGTGCCGCCTGGCCCGGCAGAGGGGGTCTCGGACGCAGGGCGCAGTTTGCCG  660
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  TGCGCAATCTGTGCCGCCTGGCCCGGCAGAGGGGGTCTCGGACGCAGGGCGCAGTTTGCCG  660

Qy    661  AGAGGAACTGGACAACTGGAGCTTCGTGCTCCAAAGTGCAGAGACCTGTTCTCAGCA  720
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  AGAGGAACTGGACAACTGGAGCTTCGTGCTCCAAAGTGCAGAGACCTGTTCTCAGCA  720

Qy    721  AAAAGGAGCTGCCGGGGTGCTGGAGACAGAGGTCACTACCAGTTCATCGACCAGAGCT  780
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  AAAAGGAGCTGCCGGGGTGCTGGAGACAGAGGTCACTACCAGTTCATCGACCAGAGCT  780

Qy    781  TCCAGGAGTTCTCGCGGCCACTGTCCGGAGACGGGGGTGCCCAGGACCG  840
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  TCCAGGAGTTCTCGCGGCCACTGTCCTGGGAGGACGGGGGTGCCCAGGACCG  840

Qy    841  CGGCTGGGCGGTTGGGACACTCCTGCGTGGGGACCGGCCAGCCCACCTTGGTGC  900
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  CGGCTGGGCGGTTGGGACACTCCTGCGTGGGGACCGGCCAGCCCACCTTGGTGC  900

Qy    901  TCACCACGCGCTCCTCTTCGGACTGCTTCGGACTGCGGTGTGAGCGGCGGATGCGGCGAATGCGAGCGCC  960
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TCACCACGCGCTTCCTCTTCGGACTGCTTCGGACTGCGGTGTGAGCGGCGGATCGGCGGACATCGAGCGCC  960

Qy    961  ACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCGGGGAC  1020
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  ACTTCGGCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCGGGGAC  1020

Qy   1021  AGGGACAGGCGCTGCCCCGAGTGCCACCAGAGTGCACCAGAGTGACCGAGGGGGCTGAGGGGGCCAAAGGGCTCGAGG  1080
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  AGGGACAGGCGCTGCCCCGAGTGCCACCAGAGTGCACCAGAGTGACCGAGGGGGCTGAGGGGGCCAAAGGGCTCGAGG  1080

Qy   1081  ACACCGAAGAGCCAGAGAGCGAGGAGGAGGAGGAGGAGCCCACTACCACCACGGAGTTGC  1140
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  ACACCGAAGAGCCAGAGAGCGAGGAGGAGGAGGAGGAGCCCACTACCACCACGGAGTTGC  1140

Qy   1141  TGTACTGCCTGTACGAGACGCAGGAGACGCCGGTTGTGCGGCCCAAGCCCTGTGCGGTTCC  1200
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  TGTACTGCCTGTACGAGACGCAGGAGACGCCGGTTGTGCGGCCCAAGCCGTGTGGGTTCC  1200

Qy   1201  CGGAGCTGGCGCGCAGGCGAGTGCGCTTCTGCCGCATGGGAGTGCGGTGTTCTTGAGTCACT  1260
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  CGGAGCTGGCGCGCAGGCGAGTGCGCTTCTGCCGCATGGGAGTGGGTGGTGTTCTTGAGTCACT  1260

Qy   1261  GCGTGAGGTGCTGCTGCGGCACTGGGCTGATCAGCTGCGAGATTGGTTGCTG  1320
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  GCGTGAGGTGCTGCTGCGGCACTGGGCTGATCAGCTGCGAGATTGGTTGCTG  1320

Qy   1321  CGGAGGAAGAGAGAGAGAGAGAGGAGCCTGGGAGGCCAGCCTCGGGTGGCGGGCA  1380
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  CGGAGGAAGAGAGAGAGAGAGAGGAGCCTGGGAGGCCAGCCTCGGGTGGCGGGCA  1380

Qy   1381  GTTCTCAAGGCACCACAAAACAACTGCCAGCCTCCTTCTTCATCCACTCTTTCAGGCAA  1440
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  GTTCTCAAGGCACCACAAAACAACTGCCAGCCTCCTTCTTCATCCACTCTTTCAGGCAA  1440

Qy   1441  TGACTGACCCACTGTGCCATCTGAGCAGCTTGTGAAGGAGAACGAGACGACGTGT  1500
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  TGACTGACCCACTGTGCCATCTGAGCAGCTTGTGAAGGAGAACGAGACGACGTGT  1500

Qy   1501  CGGTCTGCCGAGACCTTTCTGAGGCCTTCGTGTGCTGGATCGGTGCCACCGTGCTGCGCGTGGCCCAGCAGC  1560
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  CGGTCTGCCGAGACCTTTCTGAGGCCTTCGTGTGCTGGATCGGTGCCACCGTGCTGCGCGTGGCCCAGCAGC  1560

```
Qy  1561 TCCTCCACAACAGGCTCAGTGAGGCGGACTGCGTATGCTGAGTGAGGGCTAGCCGGC       1620
         |||||||||||||||||||||||||||||||||||||||||| |||||||  |||||||
Db  1561 TCCTCCACAACAGGCTCAGTGAGGCGGACTGCGTATGCTGAGTGAGGGCC-AGCCGGC      1620

Qy  1621 CGCAGTGCAGGGTGCAGACGGCTCAGGGTACAGCTGCCTGACCCCAGCGAGGCTCCAGT     1680
         ||||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||
Db  1621 CGCAGTGCAGGGTGCAGACGGTCAGGGTACAGCTGCCTGACCCCAGCGAGGCTCCAGT      1680

Qy  1681 ACCTGGTGGTATGCTTCGGCAGAGCCCGCCCCTGACCACCCTGGATCTCAGCGGCTGCC     1740
         |||||||||||||||||||||||||||||||||||||||||||||||||||||| ||||
Db  1681 ACCTGGTGGTATGCTTCGGCAGAGCCCGCCCCTGACCACCCTGGATCTCAGCGGTGCC      1740

Qy  1741 AACTGCCGCCGCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGCCAGGGATGCGGGCC  1800
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1741 AACTGCCCCCGCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGCCAGGGATGCGGGCC 1800

Qy  1801 TCAGAACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTCACAGG   1860
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1801 TGCAGAACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCACAGG  1860

Qy  1861 CCTGAAGAGGAGCAAAGCCGGACTGGTCATCACACCCAGCGCTGGACGGCCACCCAC       1920
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1861 CTGTGAAGAGGAGCAAAGCCGGATCTGGTCATCACACCCAGCGCTGGACGGCCACCCAC     1920

Qy  1921 AACCCCCAAGGAACTCATCTCGACCTTCTGAGCCTCGTGGTGGCCAGAGCAGGGTGGAAG   1980
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1921 AACCTCCCAAGGAACTCATCTCCGACCTTCTGAGCCTCGTGGTGGCCAGAGCAGGGTGGAAG  1980

Qy  1981 ACCCTAGTCAAAGTCCCTGTGGAGAGAACGAACGGCCCATTCCAAGGCACAGGAGATATTGCTC   2040
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1981 ACCCTAGTCAAAGTCCCTGTGGAGAGAACGAACGGCCCATTCCAAGGCACAGGAGATATTGCTC 2040

Qy  2041 TCGGCCTTGGGAAACTTTGAGCCGAGGAGGCCGAGAGCCGAGAGGCCCAGAC           2100
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2041 TCGGCCTTGGGAAACTTTGAGCCGAGGAGGCCGAGACAGCAGGGAGGCCCAGAC         2100

Qy  2101 ACGGCACCCTGCCCCGTCCAGGACAGGCCCAGGACCTGCCCCTCTCCCACACCTGGGGT    2160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2101 ACGGCACCCTGCCCCGTCCAGGACAGGCCCAGGACCTGCCCCTCTCCCACACCTGGGGT     2160

Qy  2161 ACCCCTTCTCCCCAGCCCACCACTACTCCACCACCTTCCTCCTCTCCTGAGACCCTCCAG    2220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2161 ACCCCTTCTCCCCAGCCCACCACTACTCCACCACCTTCCTCCTCTCCTGAGACCCTCCAG    2220

Qy  2221 CCATTCCCTTGAAAAACACCCCGACCCCAAGCCACAATAATGACAGGAGAGCTCCAA      2280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2221 CCATTCCCTTGAAAAACACCCCGACCCCAAGCCACAATAATGACAGCGAGAGCTCCAA     2280

Qy  2281 TTAACTAAGCACCTACCTGGCGGGCAGAATAACCCTTCACTGCCTGCTAATCCCCATCTGCAGT  2340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2281 TTAACTAAGCACCTACCTGGCGGGCAGAATAACCCTTCACTGCCTGCTAATCTCCCATCTGCAGT 2340

Qy  2341 GTGGCCAACAGCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGCC       2400
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2341 GTGGCCCAACAGCCCCAGAACTATGCCCACATAGACTGGAGGTAGGCAGTTCACCGCC      2400

Qy  2401 CTCCCGTGTAGGAATGAGCACCATCCCTGGCCATGGGCCCAGCCCCACAGGGCTCCAGTG    2460
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2401 CTCCCGTGTAGGAATGAGCACCATCCCTGAGCCCATGGGGTGAGCACGCCCCATGCACTGAGCACTGCTAATTCCCC 2460

Qy  2461 TGAGATCCTTTGGGAAGGGAGACTAGGGCCAGGTGGAGCAGCGCCAGGAACCCCGTGCTG    2520
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2461 TCTGAGATCTTTTGGGAAGGGAGGACTAGGGCCAGGTGGAGCAGCGGAGACCCCCGTGCTG   2520

Qy  2521 GGTGGGAAGCATGACCACATGGTGGTGAGCAGCCCCATGACTGACGGTAATTCCCC      2580
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2521 GGTGGGAAGCATGACCACATGGTGGTGAGCAGCGCCCCATGTCACTGACGGTAATTCCCC    2580

Qy  2581 TGTGGGACTCATTCTGTTCGTTCTGTTCGTTCATTACACCTGGCCAGGCGTGGTACAATACAGTCG  2640
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2581 TGTGGGACTCATTCTGTTCGTTCTGTTCGTTCATTACACCTGGCCAGGCGTGGTACAATACAGGTCG 2640

Qy  2641 GTGGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA            2689
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2641 GTGGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA          2689
```

RESULT 2
AAD08836
ID   AAD08836 standard; cDNA; 3365 BP.
XX
AC   AAD08836;
XX
DT   04-SEP-2001  (first entry)
XX
DE   Human G-protein coupled receptor-2 (GCREC-2) cDNA.
XX
KW   Human; G-protein coupled receptor-2; GCREC-2; gene therapy; cirrhosis;
KW   transgenic animal; proliferative disorder; actinic keratosis; hepatitis
KW   nephrotropic; cancer; breast; bladder; bone marrow; brain; uterus;
KW   leukaemia; adenocarcinoma; lymphoma; melanoma; myeloma; epilepsy; stroke;
KW   neurological disorder; Alzheimer's disease; Parkinson's disease; nausea;
KW   Huntington's disease; multiple sclerosis; dementia; angina pectoris;
KW   central nervous system disorder; cardiovascular disorder; hypertension;
KW   atherosclerosis; congestive heart failure; gastrointestinal disorder;
KW   dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;
KW   pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
KW   inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
KW   Addison's diseases; allergy; asthma; diabetes mellitus; antithyroid;
KW   atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
KW   rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
KW   metabolic disorder; obesity; nootropic; protozoacide; virucide; ss.
XX
OS   Homo sapiens.
XX
FH   Key          Location/Qualifiers
FT   CDS          1..2673
FT                /tag= a
FT                /product= "Human GCREC-2 protein"
XX
PN   WO200142288-A2.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US33382.
XX
PR   10-DEC-1999;  99US-0172852.
PR   22-DEC-1999;  99US-0171732.
PR   14-JAN-2000; 2000US-0176148.
PR   21-JAN-2000; 2000US-0177331.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Burford N,  Baughn MR,  Au-Young J,  Yang J,  Lu DAM,  Reddy R;
XX
DR   WPI; 2001-381635/40.
DR   P-PSDB; AAE04546.
XX
PT   New human G-protein coupled receptor polypeptides for diagnosing,
PT   preventing, and treating cell proliferative, neurological,
PT   cardiovascular, gastrointestinal, autoimmune and metabolic disorders   -
XX
PS   Claim 5; Page 159-160; 175pp; English.
XX
CC   The present sequence is human G-protein coupled receptor-2 (GCREC-2)
CC   cDNA. GCREC is useful in somatic or germline gene therapy to correct a
CC   genetic deficiency, to express a conditionally lethal gene product and
CC   to express a protein which affords protection against intracellular
CC   parasites and also for diagnosis of disorders associated with expression.
CC   of GCREC. GCREC is also useful for generating hybridisation probes useful
CC   in mapping the naturally occurring genomic sequences and to create
CC   knockin humanised animals (pigs) or transgenic animals (mice or rats) to
CC   model human diseases. GCREC is used to diagnose, prevent and treat
CC   proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
CC   hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
CC   uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)

```

```
CC   neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
CC   Parkinson's disease, multiple sclerosis, dementia and other central
CC   nervous system disorders); cardiovascular disorders (angina pectoris,
CC   hypertension, atherosclerosis, congestive heart failure);
CC   gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal
CC   spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
CC   pyrosis, pancreatitis, Crohn's disease, diarrhoea); autoimmune/
CC   inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
CC   Addison's diseases, allergies, anaemia, asthma, diabetes mellitus, atopic
CC   dermatitis, glomerulonephritis, Grave's disease, osteoarthritis,
CC   psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
CC   parasitic, protozoal and helminthic infections) and metabolic disorders
CC   (obesity, osteoporosis, viral infections).
XX
SQ   Sequence 3365 BP; 612 A; 1106 C; 1114 G; 533 T; 0 other;

  Query Match             97.7%;  Score 2626;  DB 22;  Length 3365;
  Best Local Similarity   99.8%;  Pred. No. 0;
  Matches 2629;  Conservative    0;  Mismatches    5;  Indels    0;  Gaps    0;

Qy          7 CGTGGGCGCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCGCGCCGG 66
              || |  ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        731 CGGGCACGCGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGCGGCGCGCCGG 790

Qy         67 TGCCGCAGATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGGACGAGC 126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        791 TGCCGCAGATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCGGACGAGC 850

Qy        127 TGCCGGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCG 186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        851 TGCCGGCGCTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCG 910

Qy        187 GCGCGCGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCCACGGCCCTCCTGCTGG 246
              ||||||||||||||||||||||||||| ||||||||||||||||||||||||||||||||
Db        911 GCGCGCGGTGCTAGGCGGGCTGCTGAGCAAGGCGCTGCTGCCCACGGCCCTCCTGCTGG 970

Qy        247 TGACCACGCGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGCAGTGCG 306
              ||||||||||||||||||| ||||||||||||||||||||||||||||||||||||||||
Db        971 TGACCACGCGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGCAGTGCG 1030

Qy        307 CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 366
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1031 CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 1090

Qy        367 ATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGC 426
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1091 ATGAGAGGAGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGC 1150

Qy        427 TGTGCTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGC 486
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1151 TGTGCTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGC 1210

Qy        487 TCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCATCA 546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1211 TCGGTCGGGACCTGTCGCGCACGTCCAAGACCACCACGTCAGTGTACCTGCTTTTCATCA 1270

Qy        547 CCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGGCCCCGGTTGCAGGGCGACCTGCGCA 606
              |||||||||||||||||| |||||||||||||||||||||||||||||||||||||||||
Db       1271 CCAGCGTTCTGAGCTCGGCTCCGGTAGCCGACGGCCCCGGTTGCAGGGCGACCTGCGCA 1330

Qy        607 ATCTGTGCCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGCGCAGTTTGCCGAGAAGG 666
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1331 ATCTGTGCCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGCGCAGTTTGCCGAGAAGG 1390

Qy        667 AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG 726
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1391 AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG 1450

Qy        727 AGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 786
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1451 AGCTGCCGGGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 1510

Qy        787 AGTTCCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGCTG 846
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1511 AGTTCCTCGCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGCTG 1570

Qy        847 GCGGCGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTTGGTGCTCACCA 906
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1571 GCGGCGTTGGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTTGGTGCTCACCA 1630

Qy        907 CGCGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGCCACTTCG 966
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1631 CGCGCTTCCTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGCCACTTCG 1690

Qy        967 GCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGACAGGGAC 1026
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1691 GCTGCATGGTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGACAGGGAC 1750

Qy       1027 AGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGGACACCG 1086
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1751 AGGGCTGCCCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGGACACCG 1810

Qy       1087 AAGAGCCAGAGGAGGAGGAGGAGGGAGAGGAGCCCAACTACCCACTGGAGTTGCTGTACT 1146
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1811 AAGAGCCAGAGGAGGAGGAGGAGGGAGAGGAGCCCAACTACCCACTGGAGTTGCTGTACT 1870

Qy       1147 GCCTGTACGAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGC 1206
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1871 GCCTGTACGAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGC 1930

Qy       1207 TGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGA 1266
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1931 TGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGA 1990

Qy       1267 GGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGG 1326
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       1991 GGTGCTGCCCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGG 2050

Qy       1327 AGAAGAAGAAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCAGTTCTC 1386
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2051 AGAAGAAGAAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCAGTTCTC 2110

Qy       1387 AAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAATGACTG 1446
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2111 AAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAATGACTG 2170

Qy       1447 ACCCACTGTGCCATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGTCT 1506
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2171 ACCCACTGTGCCATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGTCT 2230

Qy       1507 GCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCCGCACTGACGGAGCTGGGCCTCCTCC 1566
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2231 GCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCCGCACTGACGGAGCTGGGCCTCCTCC 2290

Qy       1567 ACAACAGGCTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCAGT 1626
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2291 ACAACAGGCTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCAGT 2350

Qy       1627 GCAGGGTGCAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGG 1686
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2351 GCAGGGTGCAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGG 2410

Qy       1687 TGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGC 1746
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2411 TGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGC 2470

Qy       1747 CCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGA 1806
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2471 CCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGA 2530

Qy       1807 CCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGTGA 1866
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       2531 CCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGTGA 2590

Qy       1867 AGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCGCTGGACGGCCACCCACAACCTC 1926
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

XX
PD   08-AUG-2002.
XX
PF   20-DEC-2001; 2001WO-US49739.
XX
PR   22-DEC-2000; 2000US-257774P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Feder J; Ramanathan C; Mintier G;
XX
DR   WPI; 2002-619252/66.
XX
PT   New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT   their fragments and homologues, useful for preventing, treating and
PT   ameliorating medical conditions, such as proliferative, gastrointestinal,
PT   or renal disorders -
XX
PS   Example 1; Page 221-222; 336pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules (I) encoding
CC   human leucine-rich repeat small intestine I (HLRRSI1) polypeptides.
CC   The nucleic acid molecules and polypeptides are useful for preventing,
CC   treating and ameliorating medical conditions, such as proliferative,
CC   gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC   related to aberrant calcium regulation or apoptosis modulation, either
CC   directly or indirectly. They are also useful for treating, preventing
CC   and/or diagnosing diseases, disorders and/or conditions of: immune system
CC   by activating or inhibiting the proliferation, differentiation, or
CC   mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC   anaemia; immunologic deficiency syndromes, e.g. human immune deficiency
CC   virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC   e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC   myasthenia gravis; asthma or allergic reactions; inflammatory
CC   conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC   e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC   ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC   Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC   pneumonia, or viral, bacterial, and fungal infections. The HLRRSI1
CC   polypeptides are useful for modulating cytokine production, antigen
CC   presentation, or other processes such as boosting immune responses.
CC   ABS63485-ABS63504 represent HLRRSI1 coding sequences and PCR primers of
CC   the invention.
XX
SQ   Sequence 2763 BP; 544 A; 890 C; 870 G; 459 T; 0 other;
```

Query Match.
Best Local Similarity   96.9%;
     97.3%;    Score 2605;   DB 24;   Length 2763;
Matches 2689;   Conservative   0;   Mismatches   0;   Indels   74;   Gaps   1;

```
Qy        1 CGGACGCGTGGGGCGGCGCCGCAGCCTGCTGACCTGATCCTGGACCAGTGCCCGACCGCGGCG   60
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db        1 CGGACGCGTGGGGCGGCGCCGCAGCCTGCTGACCTGATCCTGGACCAGTGCCCGACCGCGGCG   60

Qy       61 CGCGGGGTGCGCGACGAGTGTGGCGGAGCGCGCAGCCGGCGTGCTCTTCATCCTGGACGGCGG  120
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db       61 CGCGGGGTGCGCGACGAGTGTGGCGGAGCGCGCAGCCGGCGTGCTCTTCATCCTGGACGGCGG  120

Qy      121 ACGAGCTGCGCGGCGTGCGGGGCCGTGGGGGCCGCGGCCCGCGCGAGGCCCCGCCAGAGCCCC  180
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      121 ACGAGCTGCGCGGCGTGCGGGGCCGTGGGGGCCGCGGCCCGCGCGAGGCCCCGCCAGAGCCCC  180

Qy      181 CGAGCGGGCGCGGGGTGCTGAGTAAGGCGGTGCTGAGTAAGGCGGCTGCTGCCAGGGCCCTCC  240
            |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      181 CGAGCGGGCGCGGGGTGCTGAGTAAGGCGGTGCTGAGTAAGGCGGCTGCTGCCACGGCCCTCC  240

Qy      241 ACGAGCTGCCGGCGCTGGGCGCCGCGCGCCCCCCCGGGGAGGCTGCGAGGGCGCGCCTTCGAGGGCCG  300
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      241 ACGAGCTGCCGGCGCTGGGCGCCGCGCGCCCCCCCGGGGAGGCTGCGAGGGCGCGCCTTCGAGGGCCG  300

Qy      301 AGTGCGCCGAGGTGCGCGGCCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCT  360
            ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      301 AGTGCGCCGAGGTGCGCGGCCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCT  360
```

```
Db      2591 AGAGAGCAAAGCCGATCGGTCATCACACACCCAGCGCTGGACGGCCACCCAGCAACCTC  2650

Cy      1927 CCAAGGAACTCATCCGACCTTCTGAGGCCTCTGGGCCAGAGCAGGGTGGGAAGACCCCA  1986
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2651 CCAAGGAACTCATCTGACCTTCTGAGGCCTCTGGGCCAGAGCAGGGTGGAAGACCCCA  2710

Cy      1987 GTCAAGTCCCTGTGGAGAACGGCCATTCCAAGGGCAGGAGGATATTGCTCTCGGCC  2046
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2711 GTCAAGTCCCTGTGGAGAACGGCCATTCCAAGGGCAGGAGGATATTGCTCTCGGCC  2770

Cy      2047 TTTGGGAAACTTTGAGCCGAGAGGCCGCAGACACGGCATGTGGGAGGCCCAGACACGGCA  2106
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2771 TTTGGGAAACTTTGAGCCGAGAGGCCGCAGACACGGCATGTGGGAGGCCCAGACACGGCA  2830

Cy      2107 CCCTGCCCCGTCCAGGACAGGCCCAGACCTGCCCCTCTCCACACTGGGGTACCCCT  2166
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2831 CCCTGCCCCGTCCAGGACACGCCCAGACCTGCCCCTCTCCACACTGGGGTACCCCT  2890

Cy      2167 TCTCCCCAGCCCCACCCACTACTCCACCTTCCTCCTGGACACCCTCCAGCCATTC  2226
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2891 TCTCCCCAGCCCCACCCACTCCACCTTCCTCCTGGACACCCTCCAGCCATTC  2950

Cy      2227 CCCTGAAAACACCCCCAGCCCAAGCCACAATAATGACAGGGAGAGTCCAATAACT  2286
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      2951 CCCTTGAAAACACCCCCGACCCCCAGCCCACAATAATGACAGCGGAGTCCAATAACT  3010

Cy      2287 AAGCACCTACCTGGGGCGCGAATAACCCTTCACTGCCTGATCCCCATCTGCAGTGTGGCC  2346
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3011 AAGCACCTACCTGGGGCAGGAATAACCCTTCACTGCCTGATCCCCATCTGCAGTGTGGCC  3070

Cy      2347 CAACAGCCCCCAGAACTATGCCCACATAGACTGGAGGCAGTTCACCGTCCTCCCT  2406
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3071 CAACAGCCCCCAGAACTATGCCCACATAGACTGGAGGCAGTTCACCGTCCTCCCT  3130

Cy      2407 GTTAGGAATGAGACCATCCTGAGGCTATGGCCCAGGCCCACAGCGTCCAGTGTCTGAG  2466
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3131 GTTAGGAATGAGACCATCCTGAGGCTATGGCCCAGGCCCACAGCGTCCAGTGTCTGAG  3190

Cy      2467 ATCTTTGGGAAGGGAGGACTAGGCAGGTGAGACGCGCAGAACCCCGTGCTGGGTGG  2526
             |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3191 ATCTTTGGGAAGGGAGGACTAGGCAGGTGAGACAGCGCAGAACCCCCG GCTGGGTGG  3250

Cy      2527 AAGCATGACCACATGGTGGGTGAGCAGCCCCCATGCACTGACGTGACTGACGTGGA  2586
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3251 AAGCATGACCACATGGTGGGTGAGCAGCCCCCATGCACTGACGTGACTGACGTGGA  3310

Cy      2587 CCTCATTTCTGTTGGTTTCATTACACCTGGCCAGGCGTGTACAATACAGGTCG  2640
             |||||||||||||||||||||||||||||||||||||||||||||||||||
Db      3311 CTCATTTCTGTTGGTTTCATTACACCTGGCCAGGCGTGTGGTACAATACAGGTCG  3364
```

RESULT 3
ABS63486
ID   ABS63486 standard; cDNA; 2763 BP.
XX
AC   ABS63486;
XX
DT   15-NOV-2002 (first entry)
XX
DE   Human leucine-rich repeat small intestine I (HLRRSI1), EST #1.
XX
KW   Human; human leucine-rich repeat small intestine I; HLRRSI1; asthma;
KW   proliferative disorder; gastrointestinal disorder; renal disorder;
KW   neural disorder; reproductive disorder; calcium regulation; apoptosis;
KW   immune system; anaemia; human immune deficiency virus; HIV; cancer;
KW   blood coagulation disorder; autoimmune disorder; allergic reaction;
KW   inflammatory condition; cardiovascular disorder; ischaemia;
KW   neurological disorder; infectious disease; cytokine production;
KW   expressed sequence tag; EST; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200261086-A2.

```
Db   2521  ACAGGCGTCCAGTGCTCAGATCTTGGGAAGGGAGACTAGGGCAGTGGAGACAGGGCA  2580
           ||||||||||||||||||||| ||||||| |||||||||||||||||||||||||
Qy   2507  GAACCCCGTGCTGGGTGGGAAGCATGACCACATGGTGGGTGAGCGCCCCATGCCACTG  2566
           ||||||||||||||||||||||||||||||||||||||||||||| |||||||||||
Db   2581  GAACCCCGTGCTGGGTGGGAAGCATGACCACATGGTGGGTGAGCAGCCCCATGCACTG  2640
Qy   2567  ACGGTAAATTCCCCTGGACTCATTCTGTGGTTTTATTACACCTGGCCAGGCGTGG  2626
           ||||||||||||||||||||||||||||||||| ||||||||||||||||||||||
Db   2641  ACGGTAAATTCCCCTGGACTCATTCTGTGGTTTCTATTACACCTGGCCAGGCGTGG  2700
Qy   2627  TACAATACAGGTCGGTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2686
           |||||||||||||||||||||| ||||||||||||||||||||||||||||
Db   2701  TACAATACAGGTCGGTGCTCACAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2760
Qy   2687  AAA  2689
           |||
Db   2761  AAA  2763
```

RESULT 4
ABS63487
ID   ABS63487 standard; cDNA; 2054 BP.
XX
AC   ABS63487;
XX
DT   15-NOV-2002  (first entry)
XX
DE   Human leucine-rich repeat small intestine 1 (HLRRSI1), EST #2.
XX
KW   Human; human leucine-rich repeat small intestine I; HLRRSI1; asthma;
KW   proliferative disorder; gastrointestinal disorder; renal disorder;
KW   neural disorder; reproductive disorder; calcium regulation; apoptosis;
KW   immune system; anaemia; human immune deficiency virus; HIV; cancer;
KW   blood coagulation disorder; autoimmune disorder; allergic reaction;
KW   inflammatory condition; cardiovascular disorder; ischaemia;
KW   neurological disorder; infectious disease; cytokine production;
KW   expressed sequence tag; EST; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO2002061086-A2.
XX
PD   08-AUG-2002.
XX
PF   20-DEC-2001; 2001WO-US49739.
XX
PR   22-DEC-2000; 2000US-257774P.
XX
PA   (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI   Feder J; Ramanathan C; Mintier G;
XX
DR   WPI; 2002-619252/66.
XX
PT   New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT   their fragments and homologues, useful for preventing, treating and
PT   ameliorating medical conditions, such as proliferative, gastrointestinal,
PT   or renal disorders -
XX
PS   Example 1; Page 222-224; 336pp; English.
XX
CC   The invention relates to isolated nucleic acid molecules (I) encoding
CC   human leucine-rich repeat small intestine I (HLRRSI1) polypeptides.
CC   The nucleic acid molecules and polypeptides are useful for preventing,
CC   treating and ameliorating medical conditions, such as proliferative,
CC   gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC   related to aberrant calcium regulation or apoptosis modulation, either
CC   directly or indirectly. They are also useful for treating, preventing
CC   and/or diagnosing diseases, disorders and/or conditions of: immune system
CC   by activating or inhibiting the proliferation, differentiation, or
CC   mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC   anaemia; immunologic deficiency syndromes, e.g. human immune deficiency

CC   virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC   e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC   myasthenia gravis; asthma or allergic reactions; inflammatory
CC   conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC   e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC   ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC   Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC   pneumonia, or viral, bacterial, and fungal infections. The HLRRSI1
CC   polypeptides are useful for modulating cytokine production, antigen
CC   presentation, or other processes such as boosting immune responses.
CC   ABS63485-ABS63504 represent HLRRSI1 coding sequences and PCR primers of
CC   the invention.
XX
SQ   Sequence 2054 BP; 473 A; 631 C; 607 G; 343 T; 0 other;

Query Match             75.1%;  Score 2020.2;  DB 24;  Length 2054;
Best Local Similarity   99.9%;  Pred. No. 0;
Matches 2022;  Conservative    0;  Mismatches    0;  Indels    3;  Gaps    0;

```
Qy    665  GGAACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCGTGTTCTCAGCAAAAA   724
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1  GGAGCTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCGTGTTCTCAGCAAAAA    60

Qy    725  GGAGCTGCCGGGCGTGCTGGAGACAGAGGTCACCACCAGTTCATCGACCAGCAGCTTCCA   784
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     61  GGAGCTGCCGGGCGTGCTGGAGACAGAGGTCACCACCAGTTCATCGACCAGCAGCTTCCA   120

Qy    785  GGAGTTCCTCGGGCACTGTCCTTCTCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGC   844
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    121  GGAGTTCCTCGGGCACTGTCCTTCTCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGC   180

Qy    845  TGGCGGCGTTGGGACACTCCTCGCGTGGGGACGCGCAGCCGCACAGCCACTTGGTGCTCAC  904
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    181  TGGCGGCGTTGGGACACTCCTCGCGTGGGGACGCGCAGCCGCACAGCCACTTGGTGCTCAC  240

Qy    905  CACGGCGCTTCCTCCTTCGGACTGCTGAGCGCGGAGGATGCGCGAGCATCGAGCGGCCACTT  964
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    241  CACGGCGCTTCCTCCTTCGGACTGCTGAGCGCGGAGGATGCGCGAGCATCGAGCGGCCACTT  300

Qy    965  CGGCTGCATGGTTTCAGAGCGTGTGAAGCAGCAGGAGCCCTCGCGTGGTGGTGCAGGGACAGGG  1024
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    301  CGGCTGCATGGTTTCAGAGCGTGTGAAGCAGCAGGAGCCCTCGCGTGGTGGTGCAGGGACAGGG   360

Qy   1025  ACAGGGCTGCCCCGGAGTGGCACCAGAGTGACCAGAGGGGCCAAAGGGCTCGAGGGACAC  1084
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    361  ACAGGGCTGCCCCGGAGTGGCACCAGAGTGACCAGAGGGGCCAAAGGGCTCGAGGGACAC   420

Qy   1085  CGAAGAGCCAGAGACGAGGAGGAGGAGGAGGAGCCCAACTACCACCCTGTGCCGTTGCTA  1144
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CGAAGAGCCAGAGACGAGGAGGAGGAGGAGGAGCCCAACTACCACCCTGTGCCGTTGCTA   480

Qy   1145  CTGCCTGTACGAGACGCAGCAGAGGACGCGTTGTGCGCCAAGCCCTGTGCGGTTCCCGGA  1204
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  CTGCCTGTACGAGACGCAGCAGAGGACGCGTTGTGCGCCAAGCCCTGTGCGGTTCCCGGA   540

Qy   1205  GCTGGCGCTGCAGCGAGTGCGCTTCTGCCGCATGGACGTGGCGTGTTCTGAGCTAC_GCGT  1264
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  GCTGGCGCTGCAGCGAGTGCGCTTCTGCCGCGTGGCGTGGTTCTGAGCTAC_GCGT   600

Qy   1265  GAGGTGCTGCCCTGCTGCAGCAGGCACTGCGCGCTGATCAGCTGCAGATTGGTTGCTGCGCA  1324
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GAGGTGCTGCCCTGCTGCAGCAGGCACTGCGCGCTGATCAGCTGCAGATTGGTTGCTGCGCA   660

Qy   1325  GGAGAAGAAGAAGAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCAGCTGGGTGCGGCAGTTC  1384
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  GGAGAAGAAGAAGAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCAGCTGGGGTGCGGCAGTTC   720

Qy   1385  TCAAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAATGAC  1444
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  TCAAGGCACCACAAAACAACTGCCAGCCTCCCTTCTTCATCCACTCTTTCAGGCAATGAC   780

Qy   1445  TGACCACTGTGCCATCTGAGCAGTGCCAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGT  1504
```

Db    781  TGACCCACTGTGCCATCTGACCAGCCTCAGCGTCTCCACTGCAAACTCCCTGACGCGGT  840

Qy   1505  CTGCCGAGGACCTTCTGAGGCCTGAGGGCAGCCCCCGGCATGACGGACTGGCGCCTCC  1564
           ||||  ||||||||||||||||| ||||| |||| ||||| || ||||||||||||
Db    841  CTGCCGAGGACCTTCTGAGGCCTGAGGGCAGCCCCCGGCATGACGGAGCTGGCGCCCCT  900

Qy   1565  CCACAACAGGTCAGTGAGGCGGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCA  1624
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  CCACAACAGGTCAGTGAGGCGGGGACTGCGTATGCTGAGTGAGGGCCCAGCCTGGCGCA  960

Qy   1625  GTGCAGGGTGCAGACGGTCAGGTACACAGTTGCCTGACCCCCAGCGAGGGCTCCAGTACCT  1684
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    961  GTGCAGGGTGCAGACGGTCAGGTACACAGTTGCCTGACCCCCAGCGAGGGCTCCAGTACCT  1020

Qy   1685  GGTGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCTTCAGCGGCTGCCAACT  1744
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1021  GGTGGGTATGCTTCGGCAGAGCCCCGCCCTGACCACCTTCAGCGGCTGCCAACT  1080

Qy   1745  GCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCTGCAGCACCAGGGATGCGGCCTGCA  1804
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1081  GCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCTGCAGCACCAGGGATGCGGCCTGCA  1140

Qy   1805  GACCCTCAGTCTGGCCCTCTGTGGAGCCGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGT  1864
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1141  GACCCTCAGTCTGGCCCTCTGTGGAGCCGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGT  1200

Qy   1865  GAAGAGCAAAGCCCGGATCTGGTCATCACACACCCAGCGCTGACGGCCTGGACGGCAGCAACC  1924
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1201  GAAGAGCAAAGCCCGGATCTGGTCATCACACACCCAGCGCTGGCGGCCACCACAACC  1260

Qy   1925  TCCCAAGGAACTCATCTCGACCTTCTGAGACTCTGGTGGCCAGAGCGAGCAGGTGGAAGACCC  1984
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1261  TCCCAAGGAAC CATCTCGACCTTCTGAGGCTCTGGCCAGAGCAGGGTGGAAGACCC  1320

Qy   1985  TAGTCAAAGTCCCTGTGGAGAGAACGGCCCATTCCAAGGGCAGGAGGATATTGCTCTCGG  2044
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1321  TAGTCAAAGTCCCTGTGGAGAGAACGGCCCATTCCAAGGGCAGGAGGATATTGCTCTCGG  1380

Qy   2045  CCTTTGGAAACTTTGAGCCGGAGACAGGCCGGCATGTGGGAGGCCCGACACGG  2104
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1381  CCTTTGGAAACTTTGAGCCGGAGAGCAGGCATGTGGGAGGCCCGACACGG  1440

Qy   2105  CACCCTGCCCGTCCAGACAGGCCCAGGACCTGCCCTCTCCCACACTGGGGTACCC  2164
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1441  CACCCTGCCCGTCCAGACAGGCCCAGGACCTGCCCTCTCTCCACACTGGGGTACCC  1500

Qy   2165  CTTCTCCCCAGCCCACTACTCCACCCACCTTCCTCTCCTGTGAGACCCTCCAGCCAT  2224
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1501  CTTCTCCCCAGCCCACTACTCCACCCACCTTCCTCTCCTGTGAGACCCTCCAGCCAT  1560

Qy   2225  TCCCCTTGAAAACACCCCCGACCCCCAGCCACAATAATGCACAGCGAGAGCCCCAATTAA  2284
           |||||||||||||||||||||||||||||||||||||||||||||||||
Db   1561  TCCCCTTGAAAACACCCCCGACCCCCAGCCACAATAATGACACAGCGAGAGCTCCAATTAA  1620

Qy   2285  CTAAGCACCTACCTGGGGCCAGAATAACCTTCACTGCCTGATCCCGAGTGTGG  2344
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   1621  CTAAGCACCTACCTGGGGCAGAATAACCCTCACTGCCTGATCCCGAGTGTGG  1680

Qy   2345  CCCAACAGCCCCAGAACTATGCCCACAGTAGGCTATGGCCCAGGCCCACAGGGTCCAGTGTCG  2404
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1681  CCCAACAGCCCCAGAACTATGCCCACAGTAGGCTATGGCCCAGGCCCACAGGCGTCCAGTGTCG  1740

Qy   2405  CTGTTAGGAATGAGACCCTGGGCCAGGCCTATGGCCCAGGCCCACAGGGTCCAGTGTCTG  2464
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1741  CTGTTAGGAATGAGACCCTGGGCCAGGCCTATGGCCCAGGCCCACAGGCGTCCAGTGTCTG  1800

Qy   2465  AGATCTTTGGGAGGGAGGAGGAGGACTAGGGCAGGTGGAGCAGAGCCCGAGAACCCCGTGCTGGGTG  2524
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1801  AGATCTTTGGGAGGGAGGAGGAGGACTAGGGCAGGTGGAGCAGAGCGCGAGAACCCCGTGCTGGGTG  1860

Qy   2525  GGAAGCATGACCCATGTGGGTGAGCAGCCCATGCACTGACTGCACTGACCGTAAATTCCCCTGTG  2584
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1861  GGAAGCATGACCCACCGTGGGTGAGCAGCCCCATGCACTGACTGCACTGATGGTAAATTCCCCTGTG  1920

Qy   2585  GACTCATTTCTGTTGGTTTCTATTACACCTGGCCAGGCGGTGGTGTACAATACAGGTCGGTGC  2644
           |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1921  GACTCATTTCTGTTGGTTTCTATTACACCTGGCCAGGCGGTGGTGTACAATACAGGTCGGTGC  1980

Qy   2645  TCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2689
           |||||||||||||||||||||||||||||||||||
Db   1981  TCACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA  2025

RESULT 5
ABX97042
ID   ABX97042 standard; cDNA; 2109 BP.
XX
AC   ABX97042;
XX
DT   20-MAY-2003  (first entry)
XX
DE   Human NOV18a cDNA.
XX
KW   NOVX; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
KW   hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW   human; gene; ss.
XX
CS   Homo sapiens.
XX
PN   WO200272757-A2.
XX
PD   19-SEP-2002.
XX
PF   08-MAR-2002; 2002WO-US06908.
XX
PR   08-MAR-2001; 2001US-274101P.
PR   08-MAR-2001; 2001US-274194P.
PR   08-MAR-2001; 2001US-274281P.
PR   08-MAR-2001; 2001US-274322P.
PR   09-MAR-2001; 2001US-274849P.
PR   12-MAR-2001; 2001US-275235P.
PR   13-MAR-2001; 2001US-275578P.
PR   13-MAR-2001; 2001US-275579P.
PR   13-MAR-2001; 2001US-275601P.
PR   14-MAR-2001; 2001US-276000P.
PR   16-MAR-2001; 2001US-276776P.
PR   19-MAR-2001; 2001US-276994P.
PR   20-MAR-2001; 2001US-277239P.
PR   20-MAR-2001; 2001US-277321P.
PR   20-MAR-2001; 2001US-277327P.
PR   21-MAR-2001; 2001US-277791P.
PR   22-MAR-2001; 2001US-277833P.
PR   23-MAR-2001; 2001US-278152P.
PR   26-MAR-2001; 2001US-278894P.
PR   27-MAR-2001; 2001US-278999P.
PR   27-MAR-2001; 2001US-279036P.
PR   28-MAR-2001; 2001US-279344P.
PR   30-MAR-2001; 2001US-279338P.
PR   30-MAR-2001; 2001US-279995P.
PR   30-MAR-2001; 2001US-280233P.
PR   02-APR-2001; 2001US-280802P.
PR   02-APR-2001; 2001US-280822P.
PR   02-APR-2001; 2001US-280900P.
PR   04-APR-2001; 2001US-281194P.
PR   13-APR-2001; 2001US-283675P.
PR   30-APR-2001; 2001US-287424P.
PR   02-MAY-2001; 2001US-288066P.
PR   03-MAY-2001; 2001US-288342P.
PR   03-MAY-2001; 2001US-288528P.
PR   15-MAY-2001; 2001US-291190P.
PR   16-MAY-2001; 2001US-291099P.
PR   16-MAY-2001; 2001US-291240P.
PR   30-MAY-2001; 2001US-294885P.
PR   31-MAY-2001; 2001US-294889P.
PR   31-MAY-2001; 2001US-294999P.
PR   18-JUN-2001; 2001US-299027P.

```
PR   19-JUN-2001;  2001US-299303P.
PR   19-JUN-2001;  2001US-299310P.
PR   10-JUL-2001;  2001US-304354P.
PR   31-JUL-2001;  2001US-309199P.
PR   16-AUG-2001;  2001US-312903P.
PR   10-SEP-2001;  2001US-318462P.
PR   12-SEP-2001;  2001US-318770P.
PR   27-SEP-2001;  2001US-325430P.
PR   27-SEP-2001;  2001US-325681P.
PR   18-OCT-2001;  2001US-330380P.
PR   31-OCT-2001;  2001US-335301P.
PR   14-NOV-2001;  2001US-332172P.
PR   14-NOV-2001;  2001US-332271P.
PR   14-NOV-2001;  2001US-332272P.
PR   14-NOV-2001;  2001US-333184P.
PR   14-NOV-2001;  2001US-333272P.
PR   21-NOV-2001;  2001US-332094P.
PR   03-DEC-2001;  2001US-337426P.
PR   03-DEC-2001;  2001US-338092P.
PR   04-DEC-2001;  2001US-337185P.
PR   03-JAN-2002;  2002US-345705P.
PR   07-MAR-2002;  2002US-0092900.

XX

PA   (CURA-) CURAGEN CORP.

XX

PI   Padigaru M,   Spytek KA,   Shenoy SG,   Taupier RJ,   Pena CEA,   Li L;
PI   Zerhusen BD,   Gusev V,   Ji W,   Gorman L,   Miller CE,   Kekuda R;
PI   Paturajan M,   Gangolli E,   Vernet CAM,   Guo X,   Tcherney V;
PI   Ferrandes ER,   Casman SJ,   Malyankar UM,   Gerlach V,   Liu Y;
PI   Anderson D,   Spaderna SK,   Catterton E,   Burgess C,   Leite M,   Zhong H;
PI   Alsobrook JP,   Lepley DM,   Rieger DK.

XX

DR   WPI; 2002-723332/78.
DR   P-PSDB; ABU65075.

XX

PT   NOVX polypeptides and polynucleotides, useful for preventing or
PT   treating a disorder associated with aberrant NOVX expression or
PT   activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or
PT   bronchial asthma -

XX

PS   Claim 13; Page 149; 1103pp; English.

XX

CC   This invention describes novel human NOVX polypeptides which have
CC   cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and
CC   hypotensive activity. Pharmaceutical compositions comprising the NOVX
CC   proteins or nucleic acid molecules or NOVX antibodies are useful for
CC   preventing or treating a disorder associated with aberrant NOVX
CC   expression or activity e.g. cancer, hypertension, atherosclerosis,
CC   cardiomyopathy or bronchial asthma. The products of the invention can
CC   be used for gene therapy or in a vaccine. ABX97008-ABX97185 are cDNA
CC   fragments amplified and isolated by the PCR primers and probes
CC   represented in ABX13460-ABX13462 and ABX97186-ABX97593. ABX97008-ABX97185
CC   encode the NOVX proteins described in ABU65041-ABU65218.

XX

SQ   Sequence 2:09 BP; 371 A; 673 C; 717 G; 348 T; 0 other;

Query Match         72.1%;   Score 1938.6;   Length 2109;
Best Local Similarity  98.6%;   Pred. No. C;
Matches :972;   Conservative  C;   Mismatches  9;   Indels  18;   Gaps  3;
```

```
Db  1389  GGTGCTGCCCTGCTGCACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGGCGCAGG 1448

Qy  1327  AGAAGAAGAAGAGAGCTGGGGAAGCGGGCTCAGGCCAGCCGGGTGGGCGGCAGTTCTC 1386
          ||||||||||||||||||||||||||||||||||||| |||||
Db  1449  AGAAGAAGAAGAGAGCCTGGGGAAGCGGCTGCAGGCCAGCCTG----------------- 1492

Qy  1387  AAGGCACCACAAAAACACTGCCAGCCTCCCTCTTCTCATCCACTCTTCAGGCAATGACTG 1446
          ||||||||||||||||||||||||||||||||||||||||||||||||
Db  1493  --GGCACCACAAAACACACTGCCAGCCTCCCTCTTCATCCACTCTTTCAGGCAATGACTG 1550

Qy  1447  ACCCACTGTGCCATCTGACGCGTGTCCCACTGCAAACTCCCGACGCGGTCT 1506
          ||||||||||||||||||||||||||||||||||||| |||||||||||||
Db  1551  ACCCACTGTGCCATCTGAGCAGCCTCACGCGTGTCCCACTGCAAACTCCCGACGCGGTCC 1610

Qy  1507  GCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCGCACTGACGGAGCTGGGCTTCCCC 1566
          ||||||||||||||||||||||||||||||| |||||||||||||||||||||||||
Db  1611  GCCGAGACCTTTCTGAGGCCCTGAGGGCAGCCCCGCACTGACGGAGCTGGGCCTCCCTCC 1670

Qy  1567  ACACAGGCTCAGTGAGGCCGGACTGCGTATGCTGAGTGAGCGGGGCTAGCTGGCCGCAGT 1626
          ||||||||| ||| |||||||||||||||||||||||||||||||||||||||||||
Db  1671  ACAACAGGCTCAGTGAGGCACGAGCTCGTATGCTGAGTGAGCGGGCTCAGTACCTGG 1730

Qy  1627  GCAGGGTGCAGAGGGTCAGGGTTACAGCTGCCTGACCCCCAGCGAGGGCTTCCAGTACCTGG 1686
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1731  GCAGGGTGCAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTTCCAGTACCTGG 1790

Qy  1687  TGGGTATGCTTCGGCAGAGCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGC 1746
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1791  TGGGTATGCTTCGGCAGAGCGCCTGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGC 1850

Qy  1747  CCGCCCCATGGTGACCTACCTGTGTCGAGTCTGCAGCACCAGGGATGCGGCGCTGCAGA 1806
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1851  CCGCCCCATGGTGACCTACCTGTGTCGAGTCTGCAGCACCAGGGATGCGGCGCTGCAGA 1910

Qy  1807  CCCTCAGTCTGGCCTCTGTGGAGCTGAGGCAGCTCAGAGCGAGCAGCTCAGAGCTTCAGGCTGTGA 1866
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1911  CCCTCAGTCTGGCCTCTGTGGAGCTGAGGCGAGCAGTCACTACAGAGAGCTTCAGGCTGTGA 1970

Qy  1867  AGAGAGCAAAGCCGGATCGGTCATCAACAGCACCACGGCTGACGGCGCCACCAACCTC 1926
          |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1971  AGAGAGCAAAGCCGGATCGGTCATCAACACCACCACGCTGACGGCGCACCACAACCTC 2030

Qy  1927  CCAAGGAACTCATCTCGACCTTCTGAGGCTCGTGGTGGCCAGGAGCAGGTGGAAGACCCTA 1986
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2031  CCAAGGAACTCACTCTCGACCTTCTGAGGCTCGTGGTGGCCAGGAGCAGGTGGAAGACCCTA 2090

Qy  1987  GTCAAAGTCCCTGTGGGAGA 2006
          |||||||||||||||||||
Db  2091  GTCAAAGTCCCGTGGGAGA 2109
```

RESULT 6
AAL47132
ID    AAL47132 standard; DNA; 5556 BP.
XX
AC    AAL47132;
XX
DT    20-AUG-2002 (first entry)
XX
DE    Pyrin domain containing protein NALP6/PY6 coding sequence.
XX
KW    Pyrin domain; PYD domain; antiinflammatory; antiparkinsonian;
KW    antiarteriosclerotic; antipsoriatic; antibacterial; virucide;
KW    neuroprotective; antiarthritic; antirheumatic; antiasthmatic;
KW    nephrotropic; osteopathic; nootropic; intracellular signal transduction;
KW    inflammation; Alzheimer's disease; infection; psoriasis; asthma;
KW    arteriosclerosis; multiple sclerosis; rheumatoid arthritis; sarcoidosis;
KW    osteoarthritis; glomerulonephritis; gene; ds.
XX
OS    Unidentified.
XX
FN    WO200240668-A2.
XX

PD    23-MAY-2002.
XX
PF    30-OCT-2001; 2001WO-EP12545.
XX
PR    15-NOV-2000; 2000DE-1056687.
PR    30-NOV-2000; 2000DE-1059595.
XX
PA    (APOT-) APOTECH RES & DEV LTD.
XX
PI    Tschopp J, Martinon F;
XX
DR    WPI; 2002-427093/45.
DR    P-PSDB; AA017860.
XX
PT    New DNA encoding protein with pyrin domain, useful for treating
PT    diseases involving impaired signal transduction, particularly
PT    inflammation, also proteins and antibodies .
XX
PS    Claim 5; Fig 1; 116pp; German.
XX
CC    The present invention relates the DNA and their encoded proteins, where
CC    the proteins contain at least one PYD (pyrin) domain. These can be used
CC    to treat diseases associated with impaired intracellular signal
CC    transduction, particularly inflammation such as psoriasis,
CC    arteriosclerosis, bacterial or viral infections (particularly meningitis
CC    and pneumonia), multiple sclerosis, rheumatoid arthritis, asthma,
CC    sarcoidosis, glomerulonephritis and osteoarthritis, and also Alzheimer's
CC    and Parkinson's diseases. The present sequence is a coding sequence of
CC    the invention.
XX
SQ    Sequence 5556 BP; 914 A; 1856 C; 1879 G; 907 T; 0 other;

Query Match             58.4%;  Score 1571.2; DB 24;  Length 5556;
Best Local Similarity   89.6%;  Pred. No. 5e-272;
Matches 1772;  Conservative   0;  Mismatches   38;  Indels  168;  Gaps   1;

```
Qy      7  CGTGGGCGGCAGCCGGGCTGGCTGACCTGATCCTGACCAGTGCCCGACCAGTGCCCGCGGCGCCGCGG 66
           |||                         ||||||||||||||||||||||||||||||||||||
Db    731  CGGGCACGCGGCAGCCTGGCTGACCTGACCTGAGCCCGAGCGTGCCCCGACCGTGCCGGCGCGCCGGACCAGGC 790

Qy     67  TGCCCGAGATGCTGGCCCAGCCGCAGCGGCGGCGGCCGCGCGAGGGGCCGCGCCCTTCGAGGCGGCGGACGAGC 126
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    791  TGCCCGAGATGCTGGCCCAGCCGCAGCGGCGGCGGCCGCGCGAGGGGCCGCGCCCTTCGAGGCGGCGGACGAGC 850

Qy    127  TGCCGGCGGCGTGGGGGCGCGCGCCGCCCGAGGCGCGCCGCGCGCCCTTCGAGGCGGCGCTCCTGCTGG 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    851  TGCCGGCGGCGTGGGGGCGCGCGCCGCCCGAGGCGCGCCGCGCGCCCTTCGAGGCGGCGCTCCTGCTGG 910

Qy    187  GCGGCGCGGGTGCTAGGCGCGCGCGCGGCGCTGAGTAAGGACGGCGGCGCTGCTGCCACAGGCGGCCGG 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    911  GCGGCGCGGGTGCTAGGCGCGCGCGCGGCGCTGAGCAAGGACGGCGGCGCTGCTGCCACAGGCCGGCCGG 970

Qy    247  TGACCACGCGCGGCGGCCTGGCTGCAGGGGCGCGCGCGGTGCCAGGGGCCGCTGCAGGGGCCGCTGGTGGG 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    971  TGACCACGCGCGGCGGCCTGGCTGCAGGGGCGCGCGCGGTGCCAGGGGCCGCTGCAGGGGCCGCTGGTGGG 1030

Qy    307  CCGAGGTGCGCGGCTTCTCCGACAAGGACACAAGAGAAGAGAAGAGAAGAGAGTATTCTACAAGTTCTTCCGGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1031  CCGAGGTGCGCGGCTTCTCCGACAAGGACACAAGAGAAGAGAAGAGAAGAGAGTATTCTACAAGTTCTTCCGGG 1090

Qy    367  ATGAGAGGAGGGCCCAGCGGCGCCGCCACCGGCTTCGTGAAGGAGAACGAGACGCGTTCGCGC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1091  ATGAGAGGAGGGCCCAGCGCCGCCGCCTACCGGCTTCGTGAAGGAGAACGAGAGACGCGTGTCGCGC 1150

Qy    427  TGTGCTTCGTGCCCTTCGTGTGCGGCGATGGTGCCACCGTGCGCCCAGCAGGCTGGAGC 486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1151  TGTGCTTCGTGCCCTTCGTGTGCGGCGATGGTGCGCACCGTGCGCCAGCAGCTGGAGC 1210

Qy    487  TCGGTCGGAACCTGTCGCGGCACGTCCAAGACACCACCAGTCAGTGTACCTGCTTTTCATCA 546
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1211  TCGGTCGGGACCTGTCGCGGCACGTCCAAGACACCACCAGTCAGTGTACCTGCTTTTCATCA 1270

Qy    547  CCAGCGTTCTGAGCTGCGTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGCGGACCTGCGCA 606
```

```
Db  2351 ATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGTCTCGCCGAGACCTTT 2410
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  1519 CTGAGGCCCTGAGGGCAGCAGCCCCCGACGCGAGCCCACTGACGCGGTCTGGCCTCCCACAACAGCCTCA 1578
Db  2411 CTGAGGCCCTGAGGGCAGCAGCCCCCCCCACTGACGCGGTCTGGCCTCCTCCAAACAGCTCA 2470
Qy  1579 GTGAGGCGGGACTGCGTATGCGAGTGAGTGAGGGCCTAGCTGCGCCAGTGCAGGGTCGAGA 1638
Db  2471 GTGAGGCAGAGACTGCGTATGCGAGTGAGTGAGGGCCTAGCTGCGCCAGTGCAGGGTCGAGA 2530
Qy  1639 CGGTCAGGGTACAGTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGTATGCTTC 1698
Db  2531 CGGTCAGGGTACAGCTGCCTGACCCCCAGCCAGGGCTCCAGTACCTGGTGGTATGCTTC 2590
Qy  1699 GGCAGGACCCCGCCCCCGACCACCCTGGATCTCAGCGGCTGCCAACTGCCCGCCCCATGG 1758
Db  2591 GGCAGGACCCCGCCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGCCCGCCCCATGG 2650
Qy  1759 TGACCTACCTGTGTGCAGTCTGCAGCACCAGGATGCGGGCCTGCGAGACCCTCAGTCT 1816
Db  2651 TGACCTACCTGTGTGCAGTCTGCAGCACCAGGATGCGGGCCTGCGAGACCCTCAGCCT 2708
```

```
Db  1271 CCAGGCGGTTCTCGAGCTGGCTCCGGTAGCCAGACGGGCCCCGGTTGCGAGGGCGACCTGCGCA 1330
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy   607 ATCTGTGCGCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGGCAGTTGCGAGAAGG 666
Db  1331 ATCTGTGCGCGCCTGGCCCGCGAGGGCGTCCTCGGACGCAGGGGCCAGTTGCGAGAAGG 1390
Qy   667 AACTGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG 726
Db  1391 AACTGGAGCAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG 1450
Qy   727 AGCTGCCGGGCGGGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 786
Db  1451 AGCTGCCGGGCGGGTGCTGGAGACAGAGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG 1510
Qy   787 AGTTCCTCGGGGCACTGTCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACGCGGCTG 846
Db  1511 AGTTCCTCGCGGCACTGTCTACCTGCTGGAGGACGGCGGGGTGCCCAGGACCGCGGCTG 1570
Qy   847 GCGGGTTGGGACACTCCTGCGTGGGGACGCGCCCAGCCGCACTTGGTGCTCACCA 906
Db  1571 GCGGGCGGTTGGGACACTCCTGCGTGGGGACGCGCCCAGCCGCACTTGGTGCTCACCA 1630
Qy   907 CGGGCGGTTCCTCTCGGACCGCTCGGAGCGCGAGCGGCGGATGCGGAGCGCCACTTCG 966
Db  1631 CGGGCGGTTCCTCTCGGACGCGTGAGGGCGGAGCGGA-TGCGGGCGGACATCGAGCGGCCACTTCG 1690
Qy   967 GCTGCATGTGTTTCAGAGGGTGTGAAGCAGGAGCAGCAGGGAC 1026
Db  1691 GCTGCATGTGTTTCAGAGGGTGTGAAGCAGGAGCAGCAGGGAC 1750
Qy  1027 AGGGCTGCCCCGAGTGGCACCAGAGGTGCACCGAGGGGGCCAAAGGGCTCGAGGACACCG 1086
Db  1751 AGGGCTGCCCCGAGTGGCACCAGAGGTGCACCGAGGGGGCCAAAGGGCTCGAGGACACCG 1810
Qy  1087 AAGAGCCAGAGGAGGAGGAGGAGCCCAACTACCACCTGGAGTTGCTGTACT 1146
Db  1811 AAGAGCCAGAGGAGGAGGAGGAGCCCAACTACCACCTGGAGTTGCTGTACT 1870
Qy  1147 GCCTGTACGAGACGCAGGACGCGCGTTTGTGCGGCCAAGCCCGTGTCCCGAGC 1206
Db  1871 GCCTGTACGAGACGCAGGACGCGCGTTTGTGCGGCCAAGCCCGTGTCCCGAGC 1930
Qy  1207 TGGCGCTGCAGCGGAGTGCGCTTCTGCGCCATGGACGTGGCTCGTCTGAGCTACTGCGTGA 1266
Db  1931 TGGCGCTGCAGCGGAGTGCGCCTCTGCGCCATGGACGTGGCTCGTGCTGAGCTACTGCGTGA 1990
Qy  1267 GGTGCTGCCCCTGCTGCGGACAGGGCACTGCGGGGAAGCGGCTCCAGGACTTCTGCGCAGG 1326
Db  1991 GGTGCTGCCCCTGCTGCGGACAGGGCACTGCGGGGAAGCGGCTCCAGGACTTCTGCGCAGG 2050
Qy  1327 AGAGAAGAAGAGAGCGGCTGGGAAGCGGCTCCAGGCCTCCAGGGGCTCTC 1386
Db  2051 AGAGAAGAAGAGAGCGGCTGGGAAGCGGCTCCAGGCCTCCAGGGGCTCTGGC 2110
Qy  1387 AAGGCACCACAAAAACAACTGCCAGCCTCCCTTGTTCTTCATCCACTCTTT--------- 1433
Db  2111 TGGGGACCCAACTGGCTCAGGAAGTACCCTTTCTGACATCTGTGACATCTGCCCCCA 2170
Qy  1434 -------------------------------------------- 1434
Db  2171 CACCTCCACCAGACCCTCGGCTTCCTCCAGGGCAAGGCTTTGCCCAGAGTTCCTTTGAATA 2230
Qy  1434 -------------------------------------------- 1434
Db  2231 TAGCTCCAATTCAGCCCTGCCCCAGGCGGCTTGGCATCTGTTGAGAGGATGAATGTCACGG 2290
Qy  1434 --------------------------CAGGCAATGACTGACCCACTGTGCC 1458
Db  2291 TGTGTGGCAGGGGCTGGGCCTGGGGACCCAAAGACCCAAAAGACTGACCCCACTGTGCC 2350
Qy  1459 ATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCGGTCTCGCCGAGACCTTT 1518
```

RESULT 7
AAS00033
```
ID   AAS00033 standard; cDNA; 6461 BP.
XX
AC   AAS00033;
XX
DT   10-MAY-2001 (first entry)
XX
DE   Human ATLAS-2-encoding cDNA.
XX
KW   Human; Activated T-lymphocyte associated sequence 2; ATLAS-2; antibody;
KW   cytokine receptor; autoimmune disorder; immune disorder; cancer;
KW   T-lymphocyte-associated disorder; cell-proliferation disorder; tumour;
KW   cell differentiation disorder; immune deficiency disorder; malignancy;
KW   viral infection; bacterial infection; fungal infection; metabolism;
KW   chromosome 11p15.5; ss.
XX
OS   Homo sapiens.
XX
FH   Key             Location/Qualifiers
FT   CDS             1..5556
FT                   /*tag= a
FT                   /product= "ATLAS-2"
XX
PN   WO20011456A2.
XX
PD   01-MAR-2001.
XX
PF   19-AUG-2000; 2000WO-US22699.
XX
PR   20-AUG-1999; 99US-0150105.
PR   28-APR-2000; 2000US-...
PR   28-APR-2000; 2000US-...
PR   28-APR-2000; 2000US-0560948.
PR   28-APR-2000; 2000US-0561533.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOG ) BIOGEN INC.
XX
PI   Peyman JA, Green CD, Hsu A, Browning JA, Carulli J;
XX
DR   WPI; 2001-218453/22.
DR   P-PSDB; AAU00023.
XX
PT   New isolated activated T lymphocyte associated sequences for treating
PT   or preventing immune system associated disorders such as autoimmune
PT   disorder, immune disorder, and T-lymphocyte-associated disorder -
XX
PS   Claim 6; Fig 2; 114pp; English.
XX
```

CC    The sequence encodes human activated T-lymphocyte associated sequence 2,
CC    ATLAS-2. ATLAS-2 is related by homology to cytokine receptors and its
CC    gene is located on chromosome 11p15.5. ATLAS proteins, polynucleotides
CC    and antibodies are useful for treating/preventing conditions associated
CC    with an autoimmune disorder, immune disorder, T-lymphocyte-associated
CC    disorder, cell-proliferation disorder, cell differentiation disorder,
CC    and immune deficiency disorder and for screening for a modulator of
CC    activity or of latency or of predisposition to an immune disorder.
CC    ATLAS proteins, polynucleotides and antibodies are useful in therapeutic
CC    or prophylactic treatment of diseases associated with cell-proliferation
CC    (e.g. cancers, malignancies and tumours). The polynucleotides are useful
CC    in gene therapy, to detect ATLAS mRNA or a genetic lesion in an ATLAS
CC    gene, to modulate ATLAS activity, to screen drugs or compounds that
CC    modulate ATLAS activity or expression and to treat disorders
CC    characterised by insufficient or excessive production of ATLAS protein or
CC    production of ATLAS protein forms that have decreased or aberrant
CC    activity compared to ATLAS wild type protein and in tissue typing to
CC    identify individuals. The antibodies are useful for localisation/
CC    quantitation, isolation and detection of ATLAS and to monitor protein
CC    levels in tissue. ATLAS is useful for treating/preventing infection by
CC    bacteria, viruses and fungi, affecting bodily characteristics, e.g.
CC    biorhythms, fertility or providing analgesic effects. A host cell
CC    characteristics, and for providing analgesic effects. A host cell
CC    containing the polynucleotide is useful to produce non-human transgenic
CC    animals.
XX
SQ    Sequence 6461 BP; 1079 A; 2146 C; 2150 G; 1086 T; 0 other;

Query Match             58.4%;  Score 1571.2;  DB 22;  Length 6461;
Best Local Similarity   89.6%;  Pred. No. 5.1e-272;
Matches 1772;  Conservative    0;  Mismatches    38;  Indels  169;  Gaps    1;

Db  2441  CTGAGGCCCTGAGGGCAGCCCCCACTGACGGAGCTGGCCTCCTCCCACACAGGCTCA 2470
Qy  1579  GTGAGGCGGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCAGTGCAGGTGCAGA 1638
          ||||:||||||||||:||:|||||:||||||||||||| |||:|||||
Db  2471  GTGAGGCAGGACTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCGCAGTGCAGGTGCAGA 2530

Qy  1639  CGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGTATGCTTC 1698
          ||||:||||||||||||||||||| |||||||||||||||||||| ||||| ||||||
Db  2531  CGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGTATGCCTC 2590

Qy  1699  GGCAGAGCCCCGCCCTGACCACCCCTGGATCTCAGCGGTGCCAACTGCCGCCCCCCATGG 1759
          |||||||||||||||||||||||||||||||:|||||:||||||||||||||||||||
Db  2591  GGCAGAGCCCCGCCCTGACCACCCCTGGATCTCAGCGGTGCCAACTGCCGCCCCCCA-GG 2650

Qy  1759  TGACTACCTGTGTGCAGTCCTGCAGCACCAGGATGGGGCCTGCAGACCCTCAGTCT 1816
          ||||||||||||||||:||||:||||||||||||||||||||||||||||||||||
Db  2651  TGACTACCTGTGTGCATCCTGCAGCACCAGGATGGGGCCTGCAGACCCTCAGCCT 2708

RESULT 8
ABA08979

ID  ABA08979 standard; cDNA; 1795 BP.
XX
AC  ABA08979;
XX
DT  11-JAN-2002 (first entry)
XX
DE  Human vasopressin receptor homologue-encoding cDNA, SEQ ID NO:755.
XX
KW  Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW  haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW  inhibir; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW  proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW  myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW  chronic inflammatory condition; proliferative retinopathy;
KW  atherosclerosis; coronary heart disease; arterial ischaemia;
KW  bone disorder; osteoporosis; vascular growth disorder;
KW  tissue regeneration; wound healing; infection; immune disorder;
KW  cell culture; drug screening; gene therapy; antiinflammatory;
KW  antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW  cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW  antifungal; vulnerary; antiulcer; ss.
XX
OS  Homo sapiens.
XX
PN  WO200157188-A2.
XX
PD  09-AUG-2001.
XX
PF  05-FEB-2001; 2001WO-US03800.
XX
PR  03-FEB-2000; 2000US-0496914.
PR  27-APR-2000; 2000US-0560875.
XX
PA  (HYSE-) HYSEQ INC.
XX
PI  Tang YT; Liu C; Drmanac RT;
XX
DR  WPI; 2001-457740/49.
DR  N-PSDB; ABB11735.
XX
PT  Human proteins and DNA encoding sequences useful for preventing,
PT  treating or ameliorating a medical condition in a mammalian subject
PT  e.g. arthritis and cancer -
XX
PS  Claim 1; Page 678-679; 1963pp; English.
XX
CC  Sequences ABB:0981-ABB12330 represent 1350 novel human polypeptides, and
CC  sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC  invention also relates to vectors and recombinant host cells comprising a
CC  nucleotide of the invention, methods of producing the novel polypeptides,
CC  antibodies against the polypeptides, methods of detecting the nucleotides
CC  or polypeptides in a sample, and methods of identifying compounds which

CC  bind to polypeptides of the invention. Although novel, many of the
CC  polypeptides of the invention have homology to known proteins, thereby
CC  giving an insight into their probable biological activities, and hence
CC  potential therapeutic applications. The polypeptides of the invention may
CC  have various activities, including cytokine, cell proliferation or cell
CC  differentiation activities; stem cell growth factor activity;
CC  haematopoiesis regulatory activity; tissue growth activity;
CC  immunomodulatory activity; activin- or inhibin-related activities;
CC  chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC  thrombolytic activities; receptor or ligand activities; or may be
CC  involved in oncogenesis, cancer cell proliferation or metastasis.
CC  Depending on their biological activities, polypeptides and nucleotides of
CC  the invention are useful for preventing, treating or ameliorating medical
CC  conditions, e.g. by protein or gene therapy. Such conditions include
CC  cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC  disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC  proliferative retinopathy, atherosclerosis, coronary heart disease,
CC  arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC  vascular growth. Polypeptides involved with tissue regeneration and
CC  repair (or nucleic acids encoding them) may be used to promote wound
CC  healing (e.g., of burns, incisions and ulcers), while those with
CC  immunomodulatory activities may be used in the treatment of viral,
CC  bacterial and fungal infections in addition to immune disorders.
CC  Polypeptides with growth factor activity may be used in cell cultures to
CC  promote cell growth. For example, such polypeptides may be used to
CC  manipulate stem cells in culture to give rise to neuroepithelial cells
CC  that can be used to augment or replace cells damaged by illness,
CC  autoimmune disease or accidental damage. The polypeptides and nucleotides
CC  may also be used in the diagnosis of the above conditions, and in drug
CC  screening techniques. The present sequence represents a cDNA encoding a
CC  novel human polypeptide of the invention.
XX
SQ  Sequence 1795 BP; 297 A; 567 C; 656 G; 275 T; 0 other;

Query Match        50.9%;     Score 1368.6;   DB 22;    Length 1795;
Best Local Similarity 99.7%;   Pred. No. 7.9e-236;
Matches 1371;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

Qy      7  CGTGGGGCGCAGCTGGCCTGCTGACCTGATCCTGGACCAGTGCCCACCAGTGCCCGGCCGG 66
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    421  CGGGCACGCAGCTGGCCTGCTGACCTGATCCTGGACCAGTGCCCACCAGTGCCCGGCCGG 480

Qy     67  TGCCGCAGTGCTGGCCCAGCCCAGCGGCGTCTTCATCCTGGACGGCGCGGACGAGC 126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    481  TGCCGCAGTGCTGGCCCAGCCCAGCGGCGTCTTCATCCTGGACGGCGCGGACGAGC 540

Qy    127  TGCCGGCGGTGCTGGGGGGCCCGAGGGCCCGCGGCCCTGCACAGACCCCTTCGAGGGCGGAGCG 186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    541  TGCCGGCGGTGCTGGGGGGCCCGAGGGCCCGCGGCCCTGCACAGACCCCTTCGAGGGCGGAGGG 600

Qy    187  GCCGGCGGGGTGCTAGGGCGGTGCTGAGTAAGGCGGTGCTGAGCGAGGCGTGGAGCGTGGAGC 246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    601  GCGCGGCGGGGTGCTAGGGCGGTGCTGAGCAAGGCGGTGCTGAGCGAGGCGTGGAGGCTGTGG 660

Qy    247  TGACCACGCGCGCGCGCGCGCGCCCGGGAGGCTGCAGGGGCGCGTGTTCCCGCAGTGCG 306
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    661  TGACCACGCGCGCGCGCGCGCGCCCGGGAGGCTGCAGGGGCGCGTGTTCCCGCAGTGCG 720

Qy    307  CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    721  CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGG 780

Qy    367  ATGAGAGGAGGGCCGAGTGCGGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGGC 426
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    781  ATGAGAGGAGGGCCGAGTGCGGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGGC 840

Qy    427  TGTGCTTCGTGCCTCGTGTGCTGGATCGTGTGCACCGTGGATCGTGCGCCCAGCAGTGGAGC 486
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    841  TGTGCTTCGTGCCTCGTGTGCTGGATCGTGTGCACCGTGGATCGTGCGCCCAGCAGTGGAGC 900

Qy    487  TCGGTCGGGACCTGCGCGCGCGACGTCCAAGACACCACCAGTCAGTGTACCTGCTTTCATCA 546
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    901  TCGGTCGGGACCTGCGCGCGCGACGTCCAAGACACCACCAGTCAGTGTACCTGCTTTCATCA 960

```
XX
OS   Synthetic.
XX
FH   Key          Location/Qualifiers
FT   CDS          97..1587
FT                /*tag= a
XX
PN   WO9305073-A1.
XX
PD   18-MAR-1993.
XX
PF   11-SEP-1992; 92WO-US07786.
XX
PR   11-SEP-1991; 91US-0758921.
XX
PA   (UYBO-) UNIV BOSTON.
XX
PI   Herrera VLM, Ruiz-Opazo N;
XX
DR   WPI; 1993-100926/12.
DR   P-PSDB; AAR33389.
XX
PT   Recombinant angiotensin II cAMP/vasopressin V2 receptors and
PT   their analogues - for treating and preventing hypertension and
PT   stroke, and for diagnosing hypertension
XX
PS   Disclosure; Fig 1; 71pp; English.
XX
CC   This sequence encodes a recombinant angiotensin II cAMP/vasopressin-
CC   v2 (AII/AVPv2) receptor polypeptide. AII and AVP receptors are both
CC   G protein-coupled receptors with diverse physiological roles. AII
CC   receptors respond to the neurotransmitter angiotensin II whilst AVP
CC   receptors respond to arginine vasopressin. Isoreceptors have been
CC   described for both AVP and AII, this is consistent with their
CC   functional diversity. The recombinant receptor of the invention may
CC   be derived from a mammal, preferably a human or a rat. This sequence
CC   may be used in the construction of vectors for the expression of the
CC   recombinant receptor. Fragments of the extracellular domain of the
CC   receptor are useful as immunogens for producing antibodies which
CC   neutralise the activity of the AII/AVPv2 receptor in vivo.
CC   (Updated on 25-MAR-2003 to correct PN field.)
XX
SQ   Sequence 2251 BP; 571 A; 543 C; 637 G; 500 T; 0 other;

Query Match      31.0%;  Score 832.8;  DB 14;  Length 2251;
Best Local Similarity   70.7%;  Pred. No. 6.2e-140;
Matches 1214;  Conservative   0;  Mismatches 482;  Indels  22;  Gaps    7;
```

```
Qy   385 GCGCCTACCGCTTCG-GAAGGAGGACAGGACGACGGTCGTTCGCGTGCTCCTGCGCCCTTCG  444
             |||  ||||| |||   ||||||||||  ||||||  ||||||  ||| |||  ||||||||
Db     2 GCACTTACCGCTTCGTGAAAGAGGACAGAAGAATGAGACGCTGTACGCAC-GTGCTTTG-GCCGTTCG   61

Qy   445 TGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGTGAGCTGCGGGACCTGTCGC  504
             |||||||| ||||||||  ||||| || ||||| | |||||   ||||||||
Db    62 TGTGCTGGATCGTGTGCACCGTGCTGCTGCAGCAAATGGAGCTGGGCCGGGATCTG-TC   121

Qy   505 GCACCTCCAAGACCACCACCACTACATCTGTG-ACCTGCTCTTCATCACCAGCATGCTGA---AGT  564
             |||||||||||||||| ||||||||||| ||||||||| |||||||| || ||| | |||
Db   122 GTACCTCCAAGACCACCACTACATCTGTG·ACCTGCTCTTCATCACCAGCATGCTGA---AGT  179

Qy   565 CTCCGTAGCCGACGACGGGCCCCGGTTGCGAGGGCGACCTGGCCAATCTGTCGCCGCTGGCCC  624
             |||  ||||  ||  ||||| || |||||||||||||  ||  |||||||||||||||||||
Db   179 CTGCAGGTACCAATGACCCCGGTTCAGGGACGACAACCAGCTGCCAGGAGTGCTAG   238

Qy   625 GCGCGAGGCGCGTCCTCGGACGACGGGCAGGGCGCAGTTGCGAGAAGGACGGGAGCAACTGGAGC  684
             ||||| | ||  || ||||| || |||||||  ||  ||||||||||||||||||||||||||
Db   239 GGGAGGGCATCCTGAAGCATCAAGCACAGTTCTCAGAAAAGGACCTGGAGAATTGAAGC   298

Qy   685 TTCGTGGCTCCAAAGTGCAGACACCTGTTTCTCAGCAAAAAGGAGCTGCCGGGCGTGCTGG  744
             |||| |||||||||||||||  || | |  ||||||||  ||  |||||||||| ||||
Db   299 TTCAGGGTTCCCAAGTTCAGACAAATGTTTCTCAGCAAGAACCAGCTGCCAGGAGTGCTAG   358

Qy   745 AGAAGAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCTGCGCGCACTGT  804
```

```
Qy   547 CCAGCGTCTTGAGCTCGGCCCCGGTAGCCGACGGGCCCGGTTGCAGGGGCGACCTGGCA  606
             ||||||| |||||||||| ||||| ||||| |||||| ||  |||||||||||||||
Db   961 CCAGCGTTCGAGCTCGGCTCCGGTAGCCGACGGGCCCGGTTGCAGGGGCGACCTGGCA  1020

Qy   607 ACTGTGTGCCCCCTGGCCCCGCCAGCGCGTCTTCGGACGCGCAGTTGCGAGGAAGG  666
             || |||  || || |||||||||||| |||||||||||||||| |||||||||||||
Db  1021 ATCTGTGCCGCCGCTGGCCCCGCCAGCGCGTCTTCGGACGCGCAGTTGCGAGGAAGG  1080

Qy   667 AACTGGAGCAACTGGAGCTTGTGGCTCCAAAGTGCAGACGCGTGTTCTCAGCAAAAAGG  726
             |||||||||||||||||||| ||||||||||||||||| ||||||| ||||||||||||
Db  1081 AACTGGAGCAACTGGAGCTGGTGGCTCCAAAGTGCAGACGACGTGTTTCTCAGCAAAAAGG  1140

Qy   727 AGCTGCCGGCCGTGCTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG  786
             |||||||||| ||||||||||||||||||||||||||||||||||||| ||||||||||||||
Db  1141 AGCTGCCGGCCGTGCTGCTGGAGACAGAGGTCACCTACCAGTTCATCTGACCAGAGCTTCCAGG  1200

Qy   787 AGTTCCTGCGCGGCACTGTCCTGCTGGAGGACGGGGGGTGCCCCAGGACCGCGGGCCG  846
             |||||||| || |||| ||||| ||||||||||||||||| |||||| ||| ||||||
Db  1201 AGTTCCTGCGCGGCACACTGTCCTGCTGGAGGACGGGCGGGTGCCCCAGGACCCGCGGGCTG  1260

Qy   847 GCGGCGTGGGACACTCCTGCGTGGGACGCGCCAGCGCCGCAGCCACTCGGTGTCACCA  906
             |||||| |||||| || |||| ||||||||||  |||||| |||||||| |||||| ||
Db  1261 GCGGGGTGGGACACTCCCGTGGGACGCGCCAGCGCCGCACAGCCACTTGGTGCTCACCA  1320

Qy   907 CGGCTTCCCTTCGGACTGCTCGAGCGCGGAGCGCGGAGCGGATGCGCGACATCGAGCGCCACTTCG  966
             || ||||| |||||| || || || ||||||| |||||  ||||||||  |||||||||| ||||
Db  1321 CGGCTTCCTTCGGACTGCTCGAGCGCGGAGCGGATGCGCGACATCGAGCGCCAC·TCG  1380

Qy   967 GCTGCATGGTTCAGAGCGGTTGAAGCAGGAGGCCTGCGGTGGCAGGGACGGGACAGGGAC  1026
             ||||||||| ||||| || ||||||||||| |||||||| |||||||||||||||||||
Db  1381 GCTGCATGGTTCAGAGCGGCGTGAAGCAGGAGGCCCGGGTGCAGGGGCAGGGACAGGGAC  1440

Qy  1027 AGGCTGCCCGGAGTGACCAGAGGTGCACCAGAGGGCTCGAGGACACCG  1086
             |||||||||||||||||||||||||||||||||||||||| |||||||
Db  1441 AGGCTGCCCGGAGTGACCAGAGTGACCAGAGGGGCCAAAGGGCTCGAGGACACCG  1500

Qy  1087 AAGAGCCAGAGGAGGAGGAGGAGCAGGAGGAGCCCACTACCACTGGAGTTGCTGTACT  1146
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1501 AAGAGCCAGAGGAGGAGGAGGAGCAGGAGGAGCCCACTACCACTGGAGTTGCTGTACT  1560

Qy  1147 GCCCGTACGAGAGCGCAGGAGGACGCGGTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGC  1206
             ||| ||||||||||||||||||||||| |||||||||||||||||||||  ||||||||||
Db  1561 GCCTGTACGAGAGCGCAGGAGGACGCGGTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGC  1620

Qy  1207 TGGGCTGCCCCTGTGACAGGCACTGCGGCTTCTGCGCCATGAGACGTGGCCTGCGTTCTGAGCTACTGCGTGA  1266
             ||||||||||||| |||| |||| || || |||||| |||||||||||||| |||||||||||||||||||
Db  1621 TGGGCTGCCCCTGCTCTGTGGCACAGGCACACGCGGCTTCTGCGCCATGAGACGTGGCTGCTGTGAGCTACTGCGTGA  1680

Qy  1267 GGTGCTGGTGGACAGGCACTGCGGCTCCGCGTCATCAGCTGCAGATTGGTTGCTGCCGGCAGG  1326
             ||||||||| |||| |||||||||||| ||||| |||||| |||||| |||||||||| ||||
Db  1681 GGTGCTGGATCGTGTGCCCGCACAGGGCACACGCGGCTCCGCGTCATCAGCTGCAGATTGGTTGCTGCCGGCAGG  1740

Qy  1327 AGAAGAAGAAGAGCCCTGGGAAGCGCCTCCAGGAAGCGGCCAGCCTGGGTGGGCGGCAG  1381
             ||||||||||||||||| |||||||||||||||||||||||||||||||||||| ||
Db  1741 AGAAGAAGAAGAGCCCTGGGAAGCGGCTCCAGCGCGGCCAGCCTGGGTGGGCGGCAG  1795
```

```
RESULT 9
AAQ38723
ID   AAQ38723 standard; DNA; 2251 BP.
XX
AC   AAQ38723;
XX
DT   25-MAR-2003 (updated)
DT   15-JUL-1993 (first entry)
XX
DE   AII/AVPv2 receptor coding sequence.
XX
KW   Recombinant; angiotensin II; cAMP; vasopressin-v2; AII; AVPv2;
KW   receptor; G protein-coupled; neurotransmitter; angiotensin II; human;
KW   arginine vasopressin; isoreceptor; mammal; rat; extracellular domain;
KW   immunogen; antibody; AII/AVPv2; ss.
```

```
Db  1427 CTGTGGAGCTGACTGAGAATCCACTGAGAGAGCTTCAAGCTGTGAAGACATTAAAGCCAG 1486
         |  |||  ||  |   |||||||     ||||  ||||||  |||||||||||||||||||
Qy  1882 ATCTGGTCATCACACACCCAGCGCTGGACGCGGCCACCCAACCTTCCCAGGAACTCATCT 1941

Db  1487 ATCTGGCCATCATACATTCAAAATTGGGCACACATCCTCAGCCTCTGAAGGGATGAAGCA 1546
         ||||  || |   ||| |||    ||||| | | |  |||||  ||||||  ||   |
Qy  1942 CGACCTTCTGAGGCTCTGGTGGGCCAGAGAGGGTGGAAGACCCTAGTCAAAGTCCCTGTG 2001

Db  1547 GTGTCTTCTGA-GCCCCGGAGTCCAGAACAGGAT--ATTGCTGAAGTCAACGTTCTTCTG 1603
         |||  ||||||    |||||||||| | |||| |  |||||   ||||   ||| |||||
Qy  2002 GAGAGAACCGGCCCATTCCAAGGGCAGGAGGAGGATATTGC-TCTCGGCCTTTGGGAAACTTTT 2060

Db  1604 AATATGTCTCCCATTCAAAAGCAAGTAGATGCCAGCTTCTCCCCTCCCAGGCAAGTCTTTTC 1663
            |  |  ||||| |||   |||  |  |      |||||||||| ||||||||||||||
Qy  2061 GAGCCGGAGGGCCGCAGACAGGCATGTGGGAGCCCAG 2098

Db  1664 AAGCCAAGAAGCCACAGAAGGGCAAGCAAAGACCCAG 1701
```

RESULT 10
AAK83313
ID   AAK83313 standard; DNA; 933 BP.
XX
AC   AAK83313;
XX
DT   07-NOV-2001  (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38125.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2OC1WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179065.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184664.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216880.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.
PR   14-AUG-2000; 2000US-0225758.
PR   14-AUG-2000; 2000US-0225759.
PR   18-AUG-2000; 2000US-0226279.
PR   22-AUG-2000; 2000US-0226681.
PR   22-AUG-2000; 2000US-0226868.

```
Db   359 AAACTGTGGTCACCTACCAGTTCATTGACCAGAGCTTCCAGGAGTTCTTGGCTGCATTGT 418
Qy   805 CCTACCTGCTGGGGACAGGCGGCGGGTGTCCCAGGACCGGCGGGCTGGCCGCGTTGGGAC 864
Db   419 CATACCTACTAGACGCTGAGCGTGAGGGAGCCCCAGGAGCTCGAGGAGTGTGCAGATGCTCC 478

Qy   865 TGCGTGGGGACGCCCAGCCCACAGCCACTTGGTGCTCACCACGCGCTTCCTCTTCGGAC 924
Db   479 TGCTCTTCTGACGCGGGGCGGCTGCGTGGTCATCGGCACTCACCACTAGATTCTCTTTGGAC 538

Qy   925 TGCTGAGCGGGGAGGGGGATGCGGACATGCGGCCACTTGGCTGCATGGTTCAGAGC 984
Db   539 TGCTAAGTACAGAGAGGAGTTCGTGACATTGGAAACCATTTTGGCTGCGTGGTGCCAGGGC 598

Qy   985 GTGTGAAGCAGGAGGGCCCTGCGGTGGTGCAGGGACAGGGC---TGCCCCGGAG 1041
Db   599 GTGTGAAACAGGAACACACCTTGCGGTGGGTACAAGGACAAGGCCAACCCAAGGGTGGCGACAG 658

Qy  1042 TGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTGAGGACACCCAAGAGCCAGAGGAGG 1101
Db   659 TAGGAGCCAGAAAAAGAAGGATGGAGTCTGAAGGACGAGGAGGAAGCAGAGGAGGAGGAGG 718

Qy  1102 AGGAGGACGCGGTTGTGCGGCCAAGCCTGCTGGTACGCCCACTCGAGTTCCTGCCTACGACGAGCGC 1161
Db   719 AAGAAGAGGAGGAGGAGGAACTCAACTTGGACTGGAGCTCTCACTGTCACTGTGATGAGACAC 778

Qy  1162 AGGAGGACGCGGTTTGTGCGGCCAAGCCCTGTGCCTGGGCCTGTGCCTGCAGCGAG 1221
Db   779 AAGAGGATGATTTTGTTCGCAGGCCTCTCAGCAGCCTCCAGCAGAGATGGTACTGGAGCGAG 838

Qy  1222 TGGCTTCTCGCCGCATGGACGTGGTGTTCTGTCGTTCTGAGCTACTCGGTGACCTGCCCTGCTG 1281
Db   839 TTAGGCTGACCCGCATGGAACCTTGAGGTTCTGCACCTGCGGTGCACTGCGTGCCCCGGACG 898

Qy  1282 GACACGGCACTCGGCGGCTGATCAGCTGCAGATGTGGTTGCTGCCGCCAGGAGAAGAAGAAGA 1341
Db   899 GCCAGGCTCTGAGACTGGTGGAGCTGTGGACTGGTGGTGCGCGCAAAGGAAGAAGAAGAAGA 958

Qy  1342 GCCTGGGAAGCGCGTCCAGCCGCCAGCCTGGGTGCGCGGCAGTTCTCAAGGCACCACAAAAC 1401
Db   959 ------AGAAGAGAGCTTCATGAACCGGCTGAAGG------GTTCTCAAAGCACCGGGGAAC 1006

Qy  1402 AACTGCCAGCCGCCTCCTTCATCCAACTCTTTCAGGCAATGACTGACCCACTGTGCCATC 1461
Db  1007 AACCCCCAGCCGCCCTCCTTGCTGCCTCCACCTCTGTGAGGCAATGATTACCCAGCAATGTGGTC 1066

Qy  1462 TGAGCAGCCTCACGGTGGCTCCCACTGCGCAAACTCCCTGCCGACGCGGTCTGCCGAGGACCTTTCTG 1521
Db  1067 TGAGTATTCTGACCCTGTCACACTGCAAACTGCCCTGATGCAGTTTGTCGAGGACCTTTCTG 1126

Qy  1522 AGGCCCTGAGGGCAGCCCCCGACACTGGCGAGCTGGGCCTCTCCCACAACAGCCTCAGTG 1581
Db  1127 AGGCTCTGAGAGGTGAGTCCTTCCTAAGCAGCCTCCCTGGCCTCCTGCAACCGACTCACTG 1186

Qy  1582 AGGCGGGACTGCGTATGCTCAGTGACCTGATGGAGGGCCTAGCCTGGCCGCCAGTGCAGACGG 1641
Db  1187 AGGCAGGGCTGGCGTTTACTAAGCCAAGGGCCTGGCCTGCCCAAATGCAAGGTGCAGACAC 1246

Qy  1642 TCAGGGTGCCAGTGCTGCCCTGACCCCGGGATCTGGGCTGAGGGCTCAACTGCCCGCCCCCATGGTGA 1701
Db  1247 TCAGGATTACAGATGCCTGGGCTGGGGTCCAAGAGGTTGATCCACTACCGGTCATTGTGCCTCCAGC 1306

Qy  1702 AGAGCCCGCCCCCTGACCCTGCCAACTGCCACGGGGCCCCGCCCCCCATGGTGA 1761
Db  1307 AGAGCCCAGTGCTGCCTAACCACTCCGGACTGCCTCAGCTGCGTCAGCCTGGGTGGTGG 1366

Qy  1762 CCTACCTGTGTGCAGTGCTCAGCAGCACCAGGATGCGGCCTGCAGACCCTCAGTCTGGCCT 1821
Db  1367 AACCTCTGTTGTCAGCCCTGCGTTCAGCCCCTAAATGTGCCTAAAGACCCTCAGTCTGACTT 1426

Qy  1822 CTGTGGAGCTGAGCGCAGCACTCACTACAGAGCCTTCAGGCTGTGAAGAGACGAAAGCCCG 1881
Db  1427 ...
```

us-10-029-347-1.rng

```
PR   22-AUG-2000;   2000US-0227182.
PR   23-AUG-2000;   2000US-0227009.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   06-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230419.
PR   08-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232208.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241222.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0253990.
PR   21-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
EI   Rosen CA,   Barash SC,   Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 38125; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynuclectides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 933 BP; 146 A; 309 C; 325 G; 153 T; 0 other;

Query Match              29.7%;   Score 798.6;   DB 22;   Length 933;
Best Local Similarity    99.5%;   Pred. No. 7.3e-134;
Matches   801;   Conservative   0;   Mismatches   4;   Indels   0;   Gaps   0;

Qy      7 CGTGGGCGGCCAGCCTGGCTGACCTGATCCTGGACGTGCCCCGACCGGCGGCGGCCGG 66
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    129 CGGGCACCGCCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCCGACCGGCGGCGGCCGG 188

Qy     67 TGCCCAGATGCTCGGCCCAGCCGGCGGCTGCTCTTCATCCTGGACGCCGGACGAGC 126
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
```

Db    189  TGCCGCAGATGCTGCCCAGCCGCTGCTCTTCATCCTGGACGGCGCGCGGACGAGC  248
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    127  TGCCGGCGGTGGGGGGCCCCCGAGGCCGCGCCCTTGCACAGACCCCTTCGAGGCGGCGAGCG  186

Qy    249  TGCCGGCGGTGGGGGGCCCCCGAGGCCGCGCCCTTGCACAGACCCCTTCGAGGCGGCGAGCG  308
                 |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    187  GCGCGCGGGGTGCTAGGCGGCGGCTGCTGAGTAAGGCGCTGCTCGCCACGGCCCTTCCTGCTGG  246

Db    309  GCGCGCGGGGTGCTCTAGGCGGCGGCTGCTGAGCAAGGCGCTGCTCGCCACGGCCCTTCCTGCTGG  368
                 |||||||||||||||||||||||||||||||||||||||||||||||||
Qy    247  TGACCACGGCGCGCGCGCCGCCCCGGGAGGCTGCGAGGGCCGCCCTGTGTTCCCCGCAGTGCG  306

Qy    369  TGACCACGGCGCGCGCGCCGCCGCCCCGGGAGGCTGCGAGGGCCGCCCTGTGTTCCCCGCAGTGCG  428
                 |||||||||||||||||||||||||||||||||||||||||||||
Db    307  CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAGAAGGTATTCTACAAGTTCTTCCGGG  366

Db    429  CCGAGGTGCGCGGCTTCTCCGACAAGGACAAGAGAGAAGTATTCTACAAGTTCTTCCGGG  488
                 |||||||||||||||||||||||||||||||||||||||||||||
Qy    367  ATGAGAGGAGGGGCCGAGGGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCGTTCGCGC  426

Qy    489  ATGAGAGGAGGGGCCGAGGGCGCCCTACCGCTTCGTGAAGGAGAACGAGACGCGTTCGCGC  548
                 |||||||||||||||||||||||||||||||||||||||||||||||
Db    427  TGTGCTTCGTCGCCCTTCGTGTGCTGGATCGTGTGTGCACCCGTGCTGCGCGCCAGCGCTGGAGC  486

Db    549  TGTGCTTCGTCGCCCTTCGTGTCGTGGATCGTGTGGCACCCGTGCTGCGCGCCAGCAGCTGGAGC  608
                 |||||||||||||||||||||||||||||||||||||||||||||
Qy    487  TCGGTCGGGACCTGTCGCGCACAGACACCACCAGTGCTCAGTGTACCTGCTTTCATCA  545

Qy    609  TCGGTCGGGACCTGTCGCGCACAGACACCACCAGTGCTCAGTGTACCTGCTTTCATCA  668
                 |||||||||||||||||||||||||||||||||||||||||||||
Db    547  CCAGCGTTCTGAGCTCGGCTCCGTAGCGCAGGGCGACGGGCCCGGTTGCAGGGGCACCTGCGCA  606

Db    669  CCAGCGTTCTGAGCTCGGCTCCGTAGCGCAGGGCGACGGGCCCCGGTTGCAGGGCGACCTGCGCA  728
                 |||||||||||||||||||||||||||||||||||||||||||||
Qy    607  ATCTGTGCCGCCCTTGGCCCCGGAGGGCGTCCTCCGGACGCAGCAGCGGGCCAGTTTGCCGAGAGG  666

Qy    729  ATCTGTGCCGCCCTTGGCCCCGGAGGGCGTCCTCCGGACGCAGCAGCGGGCCAGTTTGCCGAGAAGG  788
                 |||||||||||||||||||||||||||||||||||||||||||||
Db    667  AACTGGAGCAACTGGAGCTTCGTGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG  726

Db    789  AACTGGAGCAACTGGAGCTTCGTGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGG  848
                 |||||||||||||||||||||||||||||||||||||||||||||
Qy    727  AGCTGCCCGGCGGTGCTGTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG  786

Qy    849  AGCTGCCCGGCGGTGCTGTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGG  908
                 |||||||||||||||||||||||||||||||||||||||||
Db    787  AGTTCCTGCGGCACTGTCCTACCT  811

Db    909  AGTTCCTGCGGCACTGTCCTACCT  933

RESULT 11
AAK62741
ID    AAK62741 standard; cDNA; 989 BP.
XX
AC    AAK62741;
XX
DT    06-NOV-2001  (first entry)
XX
DE    Human immune/haematopoietic antigen encoding cDNA SEQ ID NO:7801.
XX
KW    Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW    cytostatic; gene therapy; vaccine; metastasis; ss.
XX
OS    Homo sapiens.
XX
PN    WO200157182-A2.
XX
PD    09-AUG-2001.
XX
PF    17-JAN-2001; 2001WO-US01354.
XX
PR    31-JAN-2000; 2000US-0179065.

XX
PS    Claim 1; SEQ ID NO 7801; 307pp + Sequence Listing; English.
XX
CC    AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I;
CC    amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC    activity, and can be used in gene therapy and vaccine production. (I)
CC    proteins and polynucleotides may be used in the prevention, diagnosis and
CC    treatment of diseases associated with inappropriate (I) expression. For
CC    example, they may be used to treat disorders associated with decreased
CC    expression by rectifying mutations or deletions in a patient's genome
CC    that affect the activity of (I) by expressing inactive proteins or to
CC    supplement the patients own production of (I). Additionally, (I)
CC    polynucleotides may be used to produce the secreted (I) by inserting
CC    the nucleic acids into a host cell and culturing the cell to express the
CC    protein. (I) proteins and polynucleotides may be used to prevent,
CC    diagnose and treat immune/haematopoietic-related diseases, especially
CC    cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC    to AAK87694 represent human immune/haematopoietic antigen genomic
CC    sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC    represent sequences used in the exemplification of the present invention.
XX
SQ    Sequence 989 BP; 157 A; 317 C; 344 G; 164 T; 7 other;

Query Match             27.5%;    Score 740.6;    DB 22;    Length 989;
Best Local Similarity   94.5%;    Pred. No. 1.8e-123;
Matches   795;    Conservative   3;    Mismatches   39;    Indels    4;    Gaps    3;

Qy        7  CGTGGGCGCGGCAGCTGGCTGACCTGACCTGATCCTGGACCAGTGCCCGACACCGGGCGCGGGCGGCCGG    66
              ||||||||||||||||||||||||||||||||||||||||| |||||||||||||||||||||||||||
Db      129  CGGGCACGCGCAGCCTGGCTGACCTGACCTGATCCTGGACCAGTGCCCGACCACCGGGCGCGGGCGGCCGG   188

Qy       67  TGCCCAGATGCTGCGCCCAGCCGCAGCCGCTGCTTCATCCTGGACGGCGCGGACGAGC   126
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      189  TGCCCAGATGCTGCGCCCAGCCGCAGCCGCTGCTTCATCCTGGACGGCGCGGACGAGC   248

Qy      127  TGCCGGCGCTGGGGGCGCCCGAGGCGCGGCGCCGGCCCCGAGGCGCGGAGCG   186
              ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      249  TGCCGGCGCTGGGGGCGCCCGAGGCGCGGCGCCGGCCCCGAGGCGCGGAGCG   308

Qy      187  GCGCGGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCACGGCGCTGCTGGG   246
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      309  GCGCGGGGTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTGCCACGGCCCTCTGCTGG   368

Qy      247  TGACCACGCGGCGCGGGGAGGCTGCAGGGCCGCGCGTGTTCCCCCAGTGCGG   306
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      369  TGACCACGCGGCGCGGGGAGGCTGCAGGGCCGCGCGTGTTCCCCCAGTGTGCG   428

Qy      307  CCGAGGTGCGGCGGCTTCTCCGACAAGAAGAAGAAGTATTTCTACAAGTTCTTCCGGG   366
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      429  CCGAGGTGCGGCGGCTTCTCCGACAAGAACAAGAAGTATTTCTACAAGATTTCCGGG   488

Qy      367  ATGAGAGGAGGGCCGAGCGGCGCCTACGCGTTCGTGAAGGAGAACGAGACGCGTCGCGC   426
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      489  ATGAGAGGAGGGCCGAGCGGCGCCTACGCGTTCGTGAAGGAGAACGAGACGCGTTGCGC   548

Qy      427  TGTGTTCGTGCCCTTCGTGGATCGTGTGGACCGTGCTGCCCAGCAGCTGGAGC   486
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      549  TGTGTTCGTGCCCTTCGTGTGTGCTGGATCGTGTGCACCGTGCTGCCCAGCAGCTGGAGC   608

Qy      487  TCGGTCGGACCTGTCGCGCACGTCCAAGACACCAGTTCAGTGTACCTGCTTTCATCA   546
              ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      609  TCGGTCGGACCTGTCGCCCGCACGTCCAAGACACCAGTCAGTGTACCTGCTTTCATCA   668

Qy      547  CCAGGGTTCTGAGCTCGGCTCCGTAGCCGGGCGACCGGGCCCCGGTTGCAGGCGGCCA   606
              |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      669  CCACGGTTCTGAGCTCGGCTCCGTAGCCCGGTTGCAGGGCCCGGTTGCAGGCGGCCA   728

Qy      607  ATCTGTGCCCGCCTGCCCGGAGGCGTCTCCGGACGGCGCAGTTGCCGAGAAGG   666
              |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      729  ATCTGTGCCCGCCTGCCCGGAGGGGTCTCGGACGGCGCAGGCGCAARGCGGCAATTGCCGARAAGG   788

Qy      667  AACTGAGCAACTGAGCTTCGTGCTCC-AAAGTGCAGCACGG-TGTTTCAGCACAAAA   724
              ||||||||||||||||||||||||||  |||||||||||||||    ||||||||||!!

XX
SQ  Sequence 990 BP; 158 A; 318 C; 344 G; 163 T; 7 other;

Query Match           27.1%;  Score 729.6; DB 24; Length 990;
Best Local Similarity  94.4%;  Pred. No. 1.6e-121;
Matches 795; Conservative 3; Mismatches 39; Indels 5; Gaps 4;

```
Qy       7 CGTGGGCGCGGCAGCTGGTGCTGGTGACCTGATCTGGACCAGTGCCCACCGGCGCGGCCG  66
             ||||||||||||| |||||||||||||||||||||||||||||||||||||||||||||
Db     129 CGGGCACGCGGCAGCCTGGCTGACCTGATCCTGGACCAGTGCCCACCGGCGCGGCCCGG  188

Qy      67 TGCCCCAGATGCTGCCCCAGCCGCCGGCTGCTTCTTCATCCTGGACGGCGCGGACGAGAGC  126
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     189 TGCCCCAGATGCTGCCCCAGCCGCCGGCTGCTTCTTCATCCTGGACGGCGCGGACGAGAGC  248

Qy     127 TGCCGGCGCTGGGGGGCCCGGAGGGCGGCCGCCCTTGCAGACCCCTTCGAGGCGGGGAGCG  186
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     249 TGCCGGCGCTGGGGGGCCCGGAGGGCGGCCGCCCTTGCAGACCCCTTCGAGGCGGGGAGCG  308

Qy     187 GCGGCGGGGTGCTAGGCGGCTGCTGAGTAAGGCGGCTGCTGCGCCACGGCCCTCTGCTGG  246
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     309 GCGGCGGGGTGCTAGGCGGCTGCTGAGTAAGGCGGCTGCTGCGCCACGGCCCTCCTGCTGG  368

Qy     247 TGACCACGGCGCGCGCCCCGCCCCCGGGAGGCGCGCAGGGCG  306
           ||||||||||||||||||||||||||||||||||||||||||
Db     369 TGACCACGGCGCGCGCCCCGGGAGGCCGCCTGTTCCCGCAGTGCG  428

Qy     307 CCGAGGTGCGGCGGCTTCTCCGACAAGAAGGACAAGAAGAAGA:ATTCTACAAGTTCTTCCGGG  366
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     429 CCGAGGTGCGGCGGCTTCTCCGACAAGAAGGACAAGAAGAAGTAT::CTACAAGTWYTTCCGGG  488

Qy     367 ATGAGGAGGAGGGCCCAGCCGCCTACCGCTTCGTGAAGGAACGAGAGACGCTGTTCGCGC  426
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     489 ATGAGGAGGAGGGCCCAGCCGCCTACCGCTTCGTGAAGGAACGAGAGACGCTGTTCGCGC  548

Qy     427 TGTGCTTCGTGCCCCTTCGTGTGCTGGATGTGTGCACCGTGCTGCGCCAGCTGGAGC  486
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     549 TGTGTTCGTGCCCCTTCGTGTGCTGGATGTGTGCACCGTGCTGCGCCAGCTGGAGC  608

Qy     487 TCGGTGCGGACTGTCGCGGCACGTCCAAGACCACCACGTCACCTGCTTTCATCA  546
           ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     609 TCGGTGCGGACTGTCGCGGCACGTCCAAGACCACCACGTCACCTGCTTTCATCA  668

Qy     547 CCAGGCGTTCTGAGCTCGGCTTCCGGTTCCGGTTAGCCGACCGGCCCCGTTGCAGGGGACCTGCGCA  606
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     669 CCAGCGTTCTGAGCTCGGCTTCCGGTTCCGGTTAGCCGACCGGCCCCGTTGCAGGGGACCTGCGCA  728

Qy     607 ATCTGTGCCCCTGCCCCGCGAGGGCGTCCTGCGACGCAGGGCGCCAGTTGCCGGAGAGG  666
           ||||||||||||||||||||||||||||||||||| |||||||
Db     729 ATCTGTGCCCCTGCCCCGCGAGGGCGTCCTGCGACGCAAGGCGCAAGTTTGCCRAAAAGG  788

Qy     667 AACTGGAGCAACTGGAGCTTCGTGGCTCC-AAAGTGCAGACGC-TGTTTCTCAGCAAAAA  724
           |||||||||||||||||||||||||||||||||||||||||||
Db     789 AACTGGAGCAACTGGAGCTTCGTGGCTCCCAAAAGTGCAGACGCTGTTTCTCAGCAAAAA  848

Qy     725 GGAGCTG-CCGGGCGTGCTGAGACAGAGGTCACCTACAGTTCATCGACCA--GAGCTT  781
           |||||||| ||||||||||||||||||||||||||||||||||||||| ||
Db     849 GGAGCTGCCGGGCGTGCTGAGANACAGAGGTCACCTACCAGTTCATCGACCAAGAGCTT::  908

Qy     782 CCAGGAGTTCCTCGCGGCACTGCTGCTGGAGCGCGGGGTGCCCAGGACCGC  841
           ||||||| |
Db     909 NCAGGAGTTCCTCGCGGGCACTTGCTCCTACCTTGCTGGAAGGACGCGGGGTGCCCAA  968

Qy     842 GG  843
           ||
Db     969 GG  970
```

RESULT 12
ABL90582
ID  ABL90582 standard; cDNA; 990 BP.
XX
AC  ABL90582;
XX
DT  24-MAY-2002  (first entry)
XX
DE  Human polynucleotide SEQ ID NO 1144.
XX
KW  Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW  antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW  vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW  cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW  neurological disease; infection; human; secreted protein; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO200190304-A2.
XX
PD  29-NOV-2001.
XX
PF  18-MAY-2001; 2001WO-US16450.
XX
PR  19-MAY-2000; 2000US-205515P.
XX
PA  (HUMA-) HUMAN GENOME SCI INC.
XX
PI  Birse CE, Rosen CA;
XX
DR  WPI; 2002-122018/16.
DR  P-PSDB; ABB90173.
XX
PT  Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
PT  prevention of neural, immune system, muscular, reproductive,
PT  gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT  disorders -
XX
PS  Claim 4; SEQ ID NO 1144; 2081pp + Sequence Listing; English.
XX
CC  The invention relates to novel genes (AB189449-ABL90853) and proteins
CC  (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC  medical conditions e.g. by protein or gene therapy. The genes are
CC  isolated from a range of human tissues disclosed in the specification.
CC  The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC  in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC  and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC  marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC  (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC  haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC  disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC  colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC  (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC  epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC  and parasitic infections.
CC  Note: The sequence data for this patent did not form part of the
CC  printed specification, but was obtained in electronic format directly
CC  from WIPO at ftp.wipo.int/pub/published_pct_sequences.

RESULT 13
ABQ29970
ID  ABQ29970 standard; DNA; 704 BP.
XX
AC  ABQ29970;

XX
CC   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 16564.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridized to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc.; particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNPs); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ13410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 704 BP; 106 A; 82 C; 251 G; 265 T; 0 other;

Query Match       17.3%;    Score 464.6;   DB 24;   Length 704;
Best Local Similarity  78.8%;   Pred. No. 4.1e-74;
Matches  554;   Conservative    0;   Mismatches  149;   Indels    0;   Gaps    0;

Qy         50 CGACCGCGGCGCGGACGATGCTGGCCCGCAGATGCTGGCCCCAGCCGCAGCGGGCTGCTTTCATCCT 109
              ||||| |||||||||||  |||| | |||||||||| |||| |||| ||||| | || ||| |
Db          2 CGATCGCGGCGCGGCGTCGTAGATGTTGGTTAGTCGTAGCCGGTTGTTTTTTATTT 61

Qy        110 GGACGGCGCGGACGACGCTGCCGGCGCGGCGCGGGGGGGCCCCGAGCCCGCGACCACAGACCC 169
              |||||||||||||| |||| || | || |||||| || ||||  | |||||||| |||||||
Db         62 GGACGGCGCGGACGAGTTGTCGGCGTTAGGCGGGTTTGGAGGTCGCGTTTGTTAGGTTGTTT 121

Qy        170 CTTCGAGGCGGAGGCGAGCGGCGCGGCGGGTGCTAGGCGGGCTCGTCGAGTAAGGCGCGCTGCC 229
              |||||||||||||||||||||||||||||  |||||||||||||||| || |||||||| ||
Db        122 TTTCGAGGCGGCGCGAGCGGCGGCGGGTGTTAGGCGGGTTGTTGAGTAAGGCGGTTGTGTT 181

Qy        230 CACGGGCCCTCCTGCTGGTGACCACGCGCGCCCGCCGCCCCCCCGGGAGGCTGCAGGGCCGCCT 289

Qy        290 GTGTTCCCCGCAGTGCGCGCGAGGTGCGCGCCGAGGTGCGGCGCTTCTTCCGACAAGGACAAGAGAAGAAGTATTT 349
              ||||||||| ||||||||||||||| ||| || ||| ||| ||||||| ||| || |||| |||||||||| ||
Db        242 GTGTTTTTCGTAGTGGTCGTCGAGGTGCGGCGCGGTTTTTTCGATAAGGATAAGAGAAGAAGTATTT 301

Qy        350 CTACAAGTCTTCCGGATGAGAGGAGGGGCGGAGGCGGCCTACCGTCCTGTGAGGAGAA 409
              |||||||||||||||||| ||||||| ||| |||||||||||| | ||| |||||| |||
Db        302 TTATAAGTTTTTTCGGGATGAGAGGAGGGGCGGAGGCGGTTTATCGTTTCGTGAGGAGAA 361

Qy        410 CGAGACGCGTTCGCGGCTGGTTCGGTCGCGCCCTTCGTGTGCTCGGATGCGTGTGCACCGTGCT 469
              ||||||||||||||| ||| ||| || |||| ||||||| ||| || ||| |||| || |
Db        362 CGAGACGTTGTTCGCGTTGGTTCGGTCGTGTTTTTCGTGTTCGTGTGTTGCGTATCGTGTGT 421

Qy        470 GCGCCAGCAGCTGGAGCTCGGTCCGGGAGACCTGCGCGACGTCCAAGACCACCAGCGTCAGT 529
              ||| | ||||| ||||||||| || || |||| || ||||| || |||||||||| |||| |
Db        422 GCGTTAGTTAGTTGGAGTTCGGTCGGGGATTGTTGTCGTACGTTTAAGATATTACGTTAGT 481

Qy        530 GTACCTGCTTTCATCACCAGCGTCTCTGAGCTGGGTTTGAGTTCGAGTTCGGTTTGCGTTTCGGTT 589
              ||| ||||||||| | ||||||| || | |||| |||||| || |||| ||||| ||| |||||
Db        482 GTATTGTTTTTATTATTAACGGTTTTTGAGTTGGGTGAGTTCGTTTCGGTTCGTTTTCGGTT 541

Qy        590 GCAGGGCGACCTGCGCAATCTGTGCCGCACTGTCGGCCCTGGCCCGAGGGGCGCTCCTCGGACGGCAGGGC 649
              ||||||| || || ||| ||||| || | ||||||||||  ||||||||||||| ||||||| |||||||||
Db        542 GTAGGGCGATTTGCGTAATTTGTGTTGTTAATTTGTGTTTTCGGGAGGGGCGTTTGTTCGGGACGTAGGGC 601

Qy        650 GCAGTTGCCGAGAAGGAGCTGGAGCTCGGAGAGGAGCAACTGGAGCTTCGTGGCTCCAAGTGCGAGACGGCT 709
              |||||||||||||||||||||||| | || || ||||| |||| ||| ||| ||||| |||| | |||| |
Db        602 GTAGTTTGTCGAGAAGGAATTGGAGTAATTGGAGTTTCGTGGTTTTTTAAGTGTAGAGACGTT 661

Qy        710 GTTTCTCAGCAAAAAGAGAGCTGCCGGCGGTGCTGGAGACAGAG 752
              |||||| ||||||||||||||| || ||||| ||| ||||||||
Db        662 GTTTTTTAGTAAAAAGAGAGTTGTTCGGGCGGTGTTGTTGAGATAGAG 704

RESULT 24
ABQ29971/c
ID   ABQ29971 standard; DNA; 704 BP.
XX
AC   ABQ29971;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO 16562.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO200218632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
DR   WPI; 2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.

```
                    |||| ||||  ||  |||||||  ||  |||||||||   |  ||||||||  ||  ||||||||  |||||
Db     43 GTTTTTAGTAAAAGGAGTTGTCGGGCGTGTTGGAGATAGAG 1

RESULT 15
ABQ29972/c
ID   ABQ29972/c standard; DNA; 704 BP.
XX
AC   ABQ29972;
XX
DT   12-JUL-2002 (first entry)
XX
DE   Oligonucleotide for detecting cytosine methylation SEQ ID NO :6563.
XX
KW   Human; cytosine methylation; 5'-CpG-3'; uracil; cytosine; diagnosis;
KW   drug; side effect; cancer; central nervous system; cardiovascular;
KW   gastrointestinal; respiratory system; single nucleotide polymorphism;
KW   SNP; cell differentiation; ds.
XX
OS   Homo sapiens.
XX
PN   WO2002:8632-A2.
XX
PD   07-MAR-2002.
XX
PF   01-SEP-2001; 2001WO-EP10074.
XX
PR   01-SEP-2000; 2000DE-1043826.
PR   05-SEP-2000; 2000DE-1044543.
XX
PA   (EPIG-) EPIGENOMICS AG.
XX
PI   Olek A, Piepenbrock C, Berlin K, Guetig D;
XX
PT   2002-371829/40.
XX
PT   Determining the degree of cytosine methylation in genomic DNA, useful
PT   for diagnosis and prognosis, comprises selective hybridization of
PT   amplicons from chemically treated DNA -
XX
PS   Claim 12; 56pp + Sequence Listing; 56pp; German.
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 704 BP; 114 A; 82 C; 233 G; 275 T; 0 other;

       Query Match           16.1%;  Score 432.6; DB 24;  Length 704;
       Best Local Similarity 76.0%;  Pred. No. 2.2e-68;
       Matches  534;         Conservative  0;  Mismatches 169; Indels  0; Gaps  0;

Qy    49 CCGACCGCGGCGCGCGCCGGTGCGCCGAGATGCTGGCCCAGCGCGCCAGCGGCGGCTCGCTCTTCATCC 108
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   704 CCGACCGCGACGCGCCGCCGGATCCGCCGAATACGGCCAAATACTAACCCAACGACTACTCTTCATCC 645
```

---

```
XX
CC   This invention describes a novel method for determining the degree of
CC   methylation of a particular cytosine in a motif 5'-CpG-3', present in a
CC   genomic sample of DNA. The sample is treated chemically to convert
CC   cytosine (C) but not methylated C, to uracil, then part of the genomic
CC   DNA that contains the target C is amplified to form a labeled amplicon.
CC   The amplicon is hybridised to two classes, each with at least one
CC   member, of oligonucleotides and/or peptide-nucleic acid (PNA) oligomers
CC   and the degree of hybridisation to both classes is determined from the
CC   label on the amplicon. From the ratio of labels hybridised to the two
CC   classes of oligomers, the degree of methylation is calculated. The method
CC   is used: (i) for diagnosis and/or prognosis of side effects of
CC   therapeutic drugs and of a wide range of diseases, e.g. cancer, disorders
CC   of the central nervous, cardiovascular, gastrointestinal and respiratory
CC   systems etc., particularly by detecting mutations or single nucleotide
CC   polymorphisms (SNP's); and (ii) for differentiation of cell or tissue
CC   types and for investigating cell differentiation. The method allows the
CC   methylation status of many C residues to be determined simultaneously.
CC   ABQ1410-ABQ54121 represent genomic DNA sequences used to illustrate the
CC   method for determining the degree of cytosine methylation described in
CC   the disclosure of the invention.
XX
SQ   Sequence 704 BP; 265 A; 251 C; 82 G; 106 T; 0 other;

       Query Match           17.3%;  Score 464.6; DB 24;  Length 704;
       Best Local Similarity 78.8%;  Pred. No. 4.1e-74;
       Matches  554;         Conservative  0;  Mismatches 149; Indels  0; Gaps  0;

Qy    50 CGACCGCGGCGCGCGCCGGTGCCGCAGATGCTGGCCCAGCGCCGCTGCTTCATCCT 109
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   703 CGATCGCGGGCGCGCTCGGTG:CGTAGAGATGTGTTTAGTCGTAGCGGTGTTTTTA:TT 644

Qy   110 GGACGGCGGGACGAGCTGCCGGCGCTCGGGGGGCCCCGAGGCGCGCGCCCTGCACAGACCC 169
          |||||||||||||||||||||||  |||||||||||||||||||| ||-|||-||||
Db   643 GGACGCGCGGACGAGTGTGCGGCGTTGGGGGTGGGGGGTTCGGGTCGCGTT-TGTATAGA:TT 584

Qy   170 CTTCGAGGCGGCGAGCGGCGCCGGGGCGTCTAGGCGGGGCTGCTGAGTAGGCGCTGCTGCC 229
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   583 TTTCGAGGCGGCGAGCGGCGGCGCGGGGTGTTAGGCGGGT:TGGTGAGTAAGGCG:TTGGTGTT 524

Qy   230 CACGGCCCTCCTGCTGGTGACCACGGCGCGCGCCCCCCCGGGAGGCGGGGGCGCCT 289
          |||||||||||||||||||||||||||||||||||||||:||||||||||||||
Db   523 TACGGTTTTTTTGGTGATTACGCGCGTCGTCGTTT:CGGGAGGTGGTGTAGGGTCGTTT 464

Qy   290 GTGT:CCCCGCAGTGCGCCGAGGTGCGGCGGCTTCTCCGACAAGGACAAGAAGAAGTATT 349
          |||||  ||||||||||||||||||||||||||||||||||||||||||||||||
Db   463 GTGTTTTCGTAGTGCGCCAGGTGCGCGGTTTTTTCGATAAGGATAAGAAGAAGTAT:T 404

Qy   350 CTACAAGTTCTTCGGGATGAGAGGAGGAGGGGCGAGCGCGCCTACCGCTTCGTGAAGGAGAA 409
          |||| ||||||||||||||||||||||||||||||||||||||||||||||||
Db   403 TTATAAGTTTTTTCGGGATGAGAGGAGGAGGGTCGAGCGCGTTTATCGTTTCGTGAAGGAGAA 344

Qy   410 CGAGAGCGTGTTGCGCGCTGCTTCGTGCGCCCTTCGTCGTGCTGGATCGTGTGCACCGTGCT 469
          ||||||||||| || |||||||||||||| ||||||||||||||||||||
Db   343 CGAGACG:TGTTGCGGTGATTTGCGTTTCGTTTTGTGTTGGATCGTGTGTGTATCGTGTT 284

Qy   470 GCGCCAGCAGCTGAGCTCGGTCGGCGGGACCTGCGCGACGCTGCGGCCAAGCACCACGTGCAGT 529
          ||||||| |||||||||||||||||||||||||||||||||||||||||||
Db   283 GCGTTAGTAGTTGGAGTTCGGTCGGGGATTTGTTCGTCGGATTATTACGTTAGT 224

Qy   530 GTACCTGCTTTCATCACCAGCGTTCTGAGCTCGGTCCGGTAGCCGACGGGGCCCGGTT 589
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   223 GTATTTGTTTTTATTATTAGCGTTTAGCGT:TGAGTTCGGTTTCGGTTTCG:TCGGT: 164

Qy   590 GCAGGGGCGACCTGCGCAATCTGTGCCGCCTGGCCCCGGCCCGGAGGGCG:CCTCGACGACGCAGGGC 649
          ||||||||||||||||||||||||||||||||||||||||||||||
Db   163 GTAGGGCGATTGCGTTAATTGTGTCCGTTGGTTCCGTTGGTCGCGAGGGCGTTTTCGACGTAGGGC: 104

Qy   650 GCAGTTGCCGAGAAGGAACTGGAGCAACTGGAGCTTCGTGGCTGCTCCAAAGTGCGACGCT 709
          ||||| ||||||| ||||||||||||||||||||||||||||| || ||
Db   103 GTAGTTGTCGAGAAGGAATTGGAGTAATTGGAGTTTCGTGTGTT:TAAAG:GTAGACGT: 44

Qy   710 GTTTCTCAGCAAAAGGAGTTGCCGGGCGTGCTGGAGACAGAG 752
```

```
Qy    109  TGGACGGGCGCGGACGAGCTGCCGCGCTGGGGGGCCCCCAGCGCCGGCGCCTGCACAGACC  168
           |||||||||||||||||||||| ||||  |||  |||||||    |||||| ||||  ||
Cc    644  TAAACGACGCGGCAACGAACTACCGACGCTAAAAAACCCCGAAACCCGGCCCTACACAAACC  585

Qy    169  CCTTCGAGGGGGCGGCAGCGGCGGCGGGTGCTAGGCGGGTGCTGAGTAAGGCGCTGCTGC  228
           |||||||| ||||||||||||||||| ||| |||| |||||||  ||||||||||||||
Db    584  CCTTCGAAACGACGACGACGCGCGGAATACTAAACGAACTACTAAACAAAAGGCTACTAC  525

Qy    229  CCAGGGCCCTCCTGCTCGGTGCACCACGGCGGCGCCGCCCCGGGAGGCTGCAGGGCCGCC  288
           |||||||||||||  |||||||||||||||||||||||||||    |||||||||
Db    524  CCACGACCCTCCTACTAATAACCACGCGGCGCGGCGCCCCGAAAAACTACAAAACCCCC  465

Qy    289  TGTGTTCCCGCAGTGCGCCGAGGTGCGCGGCCTTCTCGACAAGGACAAGRAGAAGTATT  348
           |||||  ||||||||||||||   ||||||   ||||||||   ||| || || || ||
Db    464  CATATTCCCGCAATACGCCGAAATACGGGACTTCTCCGACAAAAACAAAAAA-ATT  405

Qy    349  TCTACAAGTTCTTCGGGATGAGAGGAGGCCGAGCGCGCCTACGCTTCGTGGAGGAGA  408
           ||||| |||| |||||||   ||||||||||||||||   |||| ||||
Db    404  TCTACAAATTCTTCCGAAATAAAAAAACCGAACGGCGCCTACCGCTTCGTAAAAAAAA  345

Qy    409  ACGAGACGCTGTTCGCGCTGTTGCTTCGTGCCCTTCGTGTGTCTGGATCGTGTGCACCGTGC  468
           ||||| || ||||||||||||    |||| |||| |||||   ||||||| |||||
Cc    344  ACGAAACGCTATTCGCGCTATTCACTTCGTACCCTTCGTATACTAAATCGTATACACCGTAC  285

Qy    469  TGCGCCAGCAGCTGGAGCTCGGTCGGGGACCTGTTCGCGCACGTCCAAGACCACCACGTCAG  528
           |||||||| |||| |||   ||| ||||||||   ||||||||||||||||||||| |||
Db    284  TACGCCAACAACTAAAAACTCGATGCAAACTATCGCGCAGTCCAAAACCACCACGTCAA  225

Qy    529  TGTACCTGCTTTTCATCACCAGCGTTCTGAGCTTCGCCGGTTCCGGTAGCCGACGGGGCCCGGT  588
           ||||||| |||||||||||||||||| ||||   ||||||||||| |||
Bb    224  TATACCTACTTTTCATCACCAACGTTCTAAACTCGACTCCGATAACCGACGACCCCGAT  165

Qy    589  TGCAGGGCGACCTGCCGCAATCTGTCCGGCCTGGCCGCGGCCGGACGCTCCTCGAGGACGCAGGG  648
           |||||||||||||||| |||| |||  ||||  || ||||||| |||||
Db    164  TACAAAAACGACCTACGCAATCTATACCCGGCCTAACCCGGAAAACGTCCTCGAACGCAAAA  105

Qy    649  CGCAGTTGCAGGAGGGAACTGGAGCAACTGGAGCTTGTGGCTCCAAAGTGCAGACGC  708
           ||||| |||  || ||||| |||||     |||| |||||||||||||
Db    104  CGCAATTTACCGACAGAAAAAAAACTAAAACAACTAAAACTTCGTAACTAAACTTCGTAACTTCGTAACTTCCAAAATACAAACGC  45

Qy    709  TGTTTCTCAGCAAAAAGGAGGCTGCCGGGGGCGTGCTGGAGACAGA  751
           |||||||||||||||||| ||||  | ||||  |||||  |
Db    44   TATTTCTCAACAAAAAAAAAACTACCGAACGTACTAAAAACAAA  2
```

Search completed: October 3, 2003, 06:57:34
Job time : 703 secs

CM protein - nucleic search, using frame_plus_p2n.model

Copyright  GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on:         October  3, 2003, 13:31:41 ; Search time 3480 Seconds
                (without alignments)
                4365.028 Million cell updates/sec

Title:          US-10-029-347-2
Perfect score:  3218
Sequence:       1 MLAQPQRLLFILDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table:  B:OSUM62
                Xgapop  10.0 , Xgapext   0.5
                Ygapop  10.0 , Ygapext   0.5
                Fgapop   6.0 , Fgapext   7.0
                Delop    6.0 , Delext    7.0

Searched:       22781392 seqs, 12552238056 residues

Total number of hits satisfying chosen parameters:   45562784

Minimum DB seq length:  0
Maximum DB seq length:  2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODE=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10029347/crunat_02102003_173436_17085/app_query.fasta_1.775
-DB=EST -QFMT=fastap -SUFFIX=rst -MINMATCH=0.1 -LOOPCL=0 -LCOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10029347_@CGN_1_1_2810 @runat_02102003_173436_17085 -NCPU=6 -ICPU=3
-NO_NMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_T:MECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=1.0 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :      EST:
                1:  em_estba.*
                2:  em_esthum.*
                3:  em_estin.*
                4:  em_estmu.*
                5:  em_estov.*
                6:  em_estpl.*
                7:  em_estro.*
                8:  em_htc.*
                9:  gb_est1.*
                10: gb_est2.*
                11: gb_htc.*
                12: gb_est3.*
                13: gb_est4.*
                14: gb_est5.*
                15: em_estfun.*
                16: em_estom.*
                17: em_gss_hum.*
                18: em_gss_inv.*
                19: em_gss_pln.*
                20: em_gss_vrt.*
                21: em_gss_mam.*
                22: em_gss_mus.*
                23: em_gss_pro.*
                24: em_gss_rod.*
                25: em_gss_phg.*
                26: em_gss_vrl.*
                27: gb_gss1.*
                28: gb_gss2.*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 2124 | 66.0 | 2730 | 11 | AK039531 | AK039531 Mus muscu |
| 2 | 927.5 | 28.8 | 798 | 9 | AW107005 | AW107005 um8d07.y |
| 3 | 901 | 28.0 | 719 | 12 | BG971110 | BG971110 602837817 |
| 4 | 830 | 25.8 | 691 | 9 | AI344276 | AI344276 tc03a07.x |
| 5 | 808 | 25.1 | 938 | 10 | BF783992 | BF783992 602107835 |
| 6 | 793 | 24.6 | 603 | 12 | BG97-489 | BG971489 602838184 |
| 7 | 731.5 | 22.7 | 832 | 12 | BG965959 | BG965959 602829753 |
| 8 | 603.5 | 18.8 | 2218 | 11 | AF054176 | AF054176 Homo sapi |
| 9 | 601.5 | 18.7 | 416 | 13 | BY262618 | BY262618 BY262618 |
| 10 | 601 | 18.7 | 670 | 10 | BB633515 | BB633515 BB633515 |
| 11 | 598.5 | 18.6 | 3368 | 11 | BC021272 | BC021272 Homo sapi |
| 12 | 598.5 | 18.6 | 3369 | 11 | BC012789 | BC012789 Homo sapi |
| 13 | 499 | 15.5 | 3400 | 11 | AK054264 | AK054264 Mus muscu |
| 14 | 487 | 15.1 | 431 | 13 | BY281505 | BY281505 BY28-505 |
| 15 | 473 | 14.7 | 3577 | 11 | AK087843 | AK087843 Mus muscu |
| 16 | 466 | 14.5 | 3218 | 11 | AK054378 | AK054378 Mus muscu |
| 17 | 458.5 | 14.2 | 3475 | 11 | AK087774 | AK087774 Mus muscu |
| 18 | 444.5 | 13.8 | 3793 | 11 | BC036506 | BC036506 Homo sapi |
| 19 (C) | 433 | 13.5 | 394 | 11 | BF561521 | BF561521 UI-R-C0-h |
| 20 | 412.5 | 12.8 | 3275 | 11 | AK054426 | AK054426 Mus muscu |
| 21 (C) | 406 | 12.6 | 488 | 9 | AI037137 | AI037137 tc90h05.y |
| 22 (C) | 390 | 12.1 | 791 | 28 | BZ129303 | BZ129303 CH230-345 |
| 23 | 381.5 | 11.9 | 1110 | 13 | BX457482 | BX457482 BX457482 |
| 24 (C) | 362.5 | 11.3 | 377 | 9 | AI558213 | AI558213 vj37a07.y |
| 25 | 357.5 | 11.1 | 939 | 11 | BX371966 | BX371966 BX371966 |
| 26 | 356 | 11.1 | 2520 | 11 | AK032446 | AK032446 Mus muscu |
| 27 | 356 | 11.1 | 4476 | 11 | AK028802 | AK028802 Mus muscu |
| 28 | 354 | 11.0 | 2394 | 11 | AK028575 | AK028575 Mus muscu |
| 29 | 352.5 | 11.0 | 664 | 29 | AG067278 | AG067278 Pan trogl |
| 30 | 339.5 | 10.6 | 921 | 13 | BQ431649 | BQ431649 AGENCOURT |
| 31 | 318 | 9.9 | 532 | 9 | AA541951 | AA541951 vj37a07.r |
| 32 | 315 | 9.8 | 4442 | 11 | AK089843 | AK089843 Mus muscu |
| 33 | 314 | 9.8 | 595 | 9 | AI883819 | AI883819 fc70d12.y |
| 34 | 313 | 9.7 | 710 | 28 | BH065379 | BH065379 RPCI-24-2 |
| 35 (C) | 310.5 | 9.6 | 3771 | 11 | BC023974 | BC023974 Homo sapi |
| 36 | 304.5 | 9.5 | 577 | 28 | AZ432279 | AZ432279 1MC217C08 |
| 37 | 297.5 | 9.2 | 791 | 10 | BG682080 | BG682080 602630057 |
| 38 | 286 | 8.9 | 934 | 10 | BE733220 | BE733220 601568455 |
| 39 | 283 | 8.8 | 629 | 14 | CB016418 | CB016418 pgnic.pk0 |
| 40 | 281 | 8.7 | 1156 | 10 | BE731237 | BE731237 6C1567129 |
| 41 | 278 | 8.6 | 2603 | 11 | AK089390 | AK089390 Mus muscu |
| 42 | 276 | 8.6 | 588 | 14 | CA559792 | CA559792 K0263E12- |
| 43 | 271.5 | 8.4 | 1114 | 14 | CD506212 | CD506212 CDA78-A06 |
| 44 | 266.5 | 8.3 | 646 | 10 | BE265950 | BE265950 601193029 |
| 45 (C) | 262 | 8.1 | 639 | 28 | AZ587775 | AZ587775 1M0395009 |

ALIGNMENTS

RESULT 1
AK039531
LOCUS       AK039531              2730 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus adult male spinal cord cDNA, RIKEN full-length
            enriched library, clone:A330055K17 product:similar to VASOPRESSIN
            RECEPTOR [Rattus norvegicus], full insert sequence.
ACCESSION   AK039531
VERSION     AK039531.1  GI:26087232
KEYWORDS    HTC; CAP trapper.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

REFERENCE   1
AUTHORS     Carninci,P. and Hayashizaki,Y.
TITLE       High-efficiency full-length cDNA cloning
JOURNAL     Meth. Enzymol. 303, 19-44 (1999)
MEDLINE     99279253
PUBMED      10349636
REFERENCE   2
AUTHORS     Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
            Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
TITLE       Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new genes
JOURNAL     Genome Res. 10 (10), 1617-1630 (2000)
MEDLINE     20449374
PUBMED      11042159
REFERENCE   3
AUTHORS     Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
            Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
            Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
            Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
            Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
            Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
            Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
TITLE       RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer
JOURNAL     Genome Res. 10 (11), 1757-1771 (2000)
MEDLINE     20530911
PUBMED      11076861
REFERENCE   4
AUTHORS     Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
            Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
            Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
            Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
            Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
            Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
            Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Tomita,M.,
            Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
            Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
            Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
            Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
            Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
            Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
            Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
            Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
            Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
            Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
            Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
            and Hayashizaki,Y.
TITLE       Functional annotation of a full-length mouse cDNA collection
JOURNAL     Nature 409 (6821), 685-690 (2001)
MEDLINE     21085660
PUBMED      11217851
REFERENCE   5
AUTHORS     The FANTOM Consortium and the RIKEN Genome Exploration Research
            Group Phase I & II Team
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
REFERENCE   6  (bases 1 to 2730)
AUTHORS     Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
            Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
            Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,T., Hirozane,C.,
            Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
            Katoh,H., Kawai,J., Kojima,Y., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
            Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,Y.,
            Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
            Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,Y., Sakazume,N.,
            Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
            Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
            Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
            Muramatsu,M. and Hayashizaki,Y.
TITLE       Direct Submission
JOURNAL     Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
            Physical and Chemical Research (RIKEN), Laboratory for Genome

            Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
            RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
            Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp/,
            URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
            Fax:81-45-503-9216)
COMMENT     cDNA library was prepared and sequenced in Mouse Genome
            Encyclopedia Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in RIKEN.
            Division of Experimental Animal Research in Riken contributed to
            prepare mouse tissues.
            Please visit our web site for further details.
            URL:http://genome.gsc.riken.go.jp/
            URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..2730
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:A330055K17"
                     /db_xref="taxon:10090"
                     /clone="A330055K17"
                     /sex="male"
                     /tissue_type="spinal cord"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="adult"
     misc_feature    1..2730
                     /note="similar to VASOPRESSIN RECEPTOR [Rattus norvegicus]
                     (SPTR|Q63035, evidence: FASTY, 82.2%ID, 100%length,
                     match:1440)"
BASE COUNT      649 a    730 c    767 g    584 t
ORIGIN

Alignment Scores:
Pred. No.:                              2730
Score:                  3.32e-193       433
Percent Similarity:     2124.00         66
Best Local Similarity:  79.71%          109
Query Match:            69.17%          18
DB:                     66.00%          8
                        11

US-10-029-347-2 (1-625) x AK039531 (1-2730)

```
Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla 20
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    233 ATACTGGCGCAACCCAATCGCTACTGTTCATCCTGGACGTGGAGACGGAGCTGCCTACT 292

Qy     21 LeuGlyGlyProGluAlaAlaProCysThrAspProProPheGluAlaAlaSerGlyAlaArg 40
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    293 TTACCATCCTCAGAGGCCACTCCCTGCAAAGACCCTTTGGAGGCTACAAGTGGCTGAGA 352

Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuAlaLeuProThrAlaLeuLeuLeuValThrThr 60
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    353 GTGCTGAGCGGGGCTGCTGAGCCAGGAGCTGCCTGCTGGTAGTGACTACA 412

Qy     61 ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysSerAlaGluVal 80
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    413 CGCCATGCGCCTACAGGAGGCTGCAGGGCAGATTGTGCTCGCCGCCACAGTGCGCAGAAATA 472

Qy     81 ArgGlyPheSerAspLysAspLysLysLysLysTyrTyrLysPhePheLysPhePheArgGluArg 100
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    473 CGCGGCTTCTCAGACAAAGACAAAAAAAAAGTATTTCTTCAAGTTCTTCCGGACGAGAGG 532

Qy    101 ArgAlaGluArgAlaTyrArgPheValLysGluAlaGluThrLeuProHisAlaLeuCysPhe 120
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    533 AAGGCAGAGCGCGCTTACCGCTTCGTGAAAGAAGCAGAGACGCTGTTTGCACTGTGCTTT 592

Qy    121 ValProPheValCysTrpIleLevalCysThrValLeuArgGlnGlnLeuGluLeuGlyArg 140
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    593 GTGCCCTTCGTGTGCTGGATCGTGTGCACAGTGCTGCAGCAGCTAGAGCTGGGCCGG 652

Qy    141 AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal 160
            :||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    653 GGTCTCTCTCGTACCTCCAAGACCACTACTTCTGTACCATG 712
```

```
Qy   515  TrpProGlnCysArgValGlnThrArgValGlnLeuProGlnArgGlyLeu  534
          ||||||:|:|||:|:|:::||:|||:|||||||||||:||||||||
Db  1760  TGGCCCAGTGCCAGGTGAAGACACTCAGGATGCAGCTGCCTGCCTTCCAAGAGAGGTGATC  1819

Qy   535  GlnTyrLeuValGlyMetLeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGly  554
          |||||||:||||||:|:|||:||:||||||||||:|||||||||:||||
Db  1820  AATTACTTGGTCATCGTGCTCCAGCAGAGCCCAGTCCTCTGGACCTCTGGACCTCAGTGGC  1879

Qy   555  CysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCys  574
          :::|||||||||||:||||||||||:||:||||||:||||||
Db  1880  TGTCAGTGCCTGGCGTCATTGTGGAACCTCTGTGTGCAGCCCTGAAGCACCCTAAATGC  1939

Qy   575  GlyLeuGlnThrLeuSerLeuAlaSerValGluSerLeuGluGlnSerLeuGlnLeuGlu  594
          :||||||:||||||:||:|||:|:|:::|||||||||||:||||||
Db  1940  TCCCTAAAGACCCTCAGTTGACCCTCTGCCGAGCTGAGTGAGAACTCACTGAGGGATCTT  1999

Qy   595  GlnAlaValLysArgAlaLysProAlaLeuValIleThrHisProAlaLeuAspGlyHis  614
          ||:||||||||:|||:||||||:||||||||::||||:||
Db  2000  CAAGCTGTGAAGACATCAAAGCCAGATCGTCCATCATATAT--TCAAAATAAGGCACA  2056

Qy   615  ProGlnProLysGlu  620
          |||:||||:|||
Db  2057  CCTCACCGGCCTCGGAAG  2074
```

RESULT 2
LOCUS       AW107005                798 bp    mRNA    linear    EST 20-OCT-1999
DEFINITION  um18d07.y1 Sugano mouse kidney mkia Mus musculus cDNA clone
            IMAGE:2192653 5' similar to TR:Q63035 Q63035 VASOPRESSIN RECEPTOR.
            , mRNA sequence.
ACCESSION   AW107005
VERSION     AW107005.1   GI:6077873
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 798)
  AUTHORS   Marra,M., Hillier,L., Kucaba,T., Martin,J., Beck,C., Wylie,T.,
            Underwood,K., Steptoe,M., Theising,B., Allen,M., Bowers,Y., Person
            ,B., Swaller,T., Gibbons,M., Pape,D., Harvey,N., Schurk,R., Ritter
            ,E., Kohn,S., Shin,T., Jackson,Y., Cardenas,M., McCann,R.,
            Waterston,R. and Wilson,R.
  TITLE     The WashU-NCI Mouse EST Project 1999
  JOURNAL   Unpublished
COMMENT     Other_ESTs: um18d07.x1
            Contact: Marra M/WashU-NCI Mouse EST Project 1999
            Washington University School of Medicine
            4444 Forest Park Parkway, Box 8501, St. Louis, MO 63108, USA
            Tel: 314 286 1800
            Fax: 314 286 1810
            Email: mouseest@watson.wustl.edu
            This clone is available royalty-free through LLNL ; contact the
            IMAGE Consortium (info@image.llnl.gov) for further information.
            MGI:1005105
            Seq primer: custom primer used
            High quality sequence stop: 503.
FEATURES             Location/Qualifiers
     source          1..798
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:2192653"
                     /sex="female"
                     /dev_stage="adult"
                     /lab_host="DH10B"
                     /clone_lib="Sugano mouse kidney mkia"
                     /note="Organ: kidney; Vector: pME18S-FL3; Site_1: DraII
                     (CACCATGTG); Site_2: DraII (CACCATGTG); 1st strand cDNA
                     was primed with an oligo(dT) primer
                     [ATGTGGCCTTTTTTTTTTTTTTT]; double-stranded cDNA was
                     ligated to a DraII adaptor [TGTTGGCCTACTGG], digested
```

```
Qy   161  LeuSerSerAlaProValAlaAspGlyArgLeuGlnGlyAspLeuArgAsnLeuCys  180
          |||:|:|||||||:||||||::|||||:||||||||||::||
Db   713  CTGAAGTCTGCAGGC---ACCAATGGACACCCGGGTTCAGGGAGGAGCTGAGAACGCTGTGT  769

Qy   181  ArgLeuAlaArgGluGlyVal-LeuGly---ArgArgAlaGlnPheAlaGluLysGluLeu  199
          |:|:|||||:|:||||||||||:||:|||||||||:||||||||
Db   770  CGCCTGGCCAGGGAGGGGGCATCCTGGATCATCATAAAGCACAATTCTCAGAAGAGAGGACTG  829

Qy   200  GluGlnLeuLeuArgGlyGluLeuThrLeuProPheLeuSerLysLysGluLeu  219
          :||:|:|||:||||||:|||:|::|:|||:|||:|:||
Db   830  GAGAAATTGAAGCTTCGGGGTTCCCAAGTTCAGACAATATTTCTCAACAAGAAAGAGTA  889

Qy   220  ProGlyValLeuGluThrTyrArgGlnPheIleAspGlnSerPheGlnGluPhe  239
          |||:|:||:||||:|:|||||:||:||||:||::|||:|||::||
Db   890  CCAGGAGTGCTAAAAACTGAGTCACCTACCAGTTCATTGACCAGAGC-TCCAGGAGTTC  949

Qy   240  LeuAlaAlaLeuSerTyrLeuLeuGluAspGlyGlyValProArgArgThrAlaAlaGlyGly  259
          ||:||||:|||||:|||::||:||||::::||:|||:|||||||
Db   950  TTGGCTGCATTGTCATACCTACTAGAAGCTGAGGAACTCCGGGACGCCCCAGGGGGT  1009

Qy   260  ValGlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrThrArg  279
          ||:|||:||:|||:||||||:||||:||||||:|||::::||||||||||
Db  1010  GTGCAGAGCTCCTGAACTCGACGGCGTACGTGGTCATCTGCACTGACCACCCCGA  1069

Qy   280  PheLeuPheGlyLeuLeuSerAlaGluArgMetArgGlyAspIleGluArgHisPheGlyCys  299
          :||:|||:|||||||::::|:||||:||::|:|||||:::|:|||
Db  1070  TTCCTCTTTGGCCTGCAAATACAGAGGGGCTTCGTGACATTGGAAAACCATTTTGGCCGT  1129

Qy   300  MetValSerGluArgValLysGlnGlnLeuGlnAlaLeuArgTrpValGlnGlnGlyV------  317
          :::||||||||:|::::|||:|:|||||||:||||::|
Db  1130  GTGGTGCCAGATCATGTGAAAAAAGGACACCCTGGTGGTACAAGGACCAGAGCCACCCC  2189

Qy   318  GlnGlyCysProGlyValAlaProGluValThrGluGlyValAlaLysGlyLeuAspThr  337
          |:|||||||||::::|||
Db  1190  AAGGGGCCCACTAGGGGCCAAAAAGACTGTGAG----------CTGGAGGACATC  1237

Qy   338  GluGluProGluGluGluGluGluLeuTyrGluGluPro---AsnTyrProLeuGlyLeuLeu  356
          |:||||||:||::||:|:::|||
Db  1238  GAGGACGAGCAGGAGGAGGAAGAGGAGGAGGAGAGGACCTCAATTTGGACTGGAGCTGCTG  1297

Qy   357  TyrCysCysLeuTyrGlnThrGluThrGluGluAspAlaPheValArgArgIleAlaLeuCysArgArgPhePro  376
          :||||||||:|||::::||::|||||||||||||||
Db  1298  TACTGCTGTCTATGAGACAAGCAGGAGGAGGATTTTGTTGCACAGGCTCTCAGCAGCCTTCCA  1357

Qy   377  GluLeuAlaLeuGlnLeuArgValArgPheCysSerArgMetAspValAlaValLeuSerTyrCys  396
          ::||:|||:||||:||||:||:||:||:::::|||||:||
Db  1358  GAGATAGTACTGAGCGAGTCAGGTTGACCCGCATGGACGTTCTGAAGTTCTGAATTACTGC  1417

Qy   397  ValArgCysCysHisProLeuAlaGlyGlnAlaLeuLeuArgLeuLeuSerGlyLeuValAlaAla  416
          :|::|:|::::||:|||:||||||:|:|:|||||||
Db  1418  GTGCAGTGCTGCTGCCCAGATGGCCAGGCTCTGAGACGTCTGAGCGTGGGCTGGTGGCGGCA  1477

Qy   417  GlnGlu-----LysLysLysSerLeuGlyLeuArgGluGlnAlaSerLeuGlyGlyGly  434
          :|||||||:||||:|:|:||||
Db  1478  AAGGAGAAGAAGAAGAAGAGAGCCTGGTGAAGCGGCTGAAG----------  1522

Qy   435  GlySerSerGlnGlyThrThrLysGlnLeuGlnLeuAlaSerLeuLeuHisProLeuPheGln  454
          ::||||:|::::||:||:||||||::||||
Db  1523  --GGCTCTCAAAGACCCAGAAACACACCCCAGTCCTCTGCTGCGCCCACTCTCTGCGAG  1579

Qy   455  AlaMetThrAspProLeuCysHisLeuSerLeuThrLeuSerHisCysLysLeuLysPro  474
          ::|||:||:|||:|::::||||:|||::|||
Db  1580  ACAATGACTACCCCGAAATGTCATCTGAGTGTTCTGATCTTGTCACACTGCACACTCCCT  1639

Qy   475  AspAlaValCysArgAspLeuSerGlyIleAlaLeuAlaLeuArgLeuAlaProAlaLeuThrGluLeu  494
          |||||:||:||||:||:||::||||||||:||:|
Db  1640  GATGCAGTTGCCAGACCTTCCGAGGCCCTGAAGG-AGC-TCCTGCCCTAAGGGAGTTG  1699

Qy   495  GlyLeuLeuHisAsnArgLeuSerGluAlaGlyLeuLeuArgMetLeuSerLeuValArgLeuGluGlyValLeuAla  514
          |||::|||:::|||:::||:||||::|:|||||||||
Db  1700  GGGCTCCTCCCCAGAGCCCGGCTCACCAACACCAGCCTGCGTTTACTGTGCGAAGGCCTTGGCT  1759
```

Qy      291  ArgAspIleGluArgHisPhe 297
             |||||||||||||||||||||
Db      775  CGTGACATTGG-AAACATTT 794

RESULT 3
BG971110
LOCUS       BG971110          719 bp      mRNA      linear    EST 12-JUN-2001
DEFINITION  602837817F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4971940
            5', mRNA sequence.
ACCESSION   BG971110
VERSION     BG971110.1  GI:14358747
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 719)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10958  row: b  column: 05
            High quality sequence start: 6
            High quality sequence stop: 717.
FEATURES             Location/Qualifiers
     source          1..719
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4971940"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Kid14"
                     /note="Organ: kidney; Vector: pCMV-SPORT6; Site_1: NotI;
                     Site_2: SalI; Cloned unidirectionally. Primer: Oligo dT.
                     Average insert size 1.75 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."
BASE COUNT   176 a   189 c   205 g   149 t
ORIGIN

Alignment Scores:
Pred. No.:                              Length:          719
Score:            2.98e-76              Matches:         186
Percent Similarity: 901.00              Conservative:     21
Best Local Similarity: 87.34%           Mismatches:       26
Query Match:      78.48%                Indels:            4
DB:               28.00%                Gaps:              2
                   12

US-10-029-347-2 (1-625) x BG971110 (1-719)

Qy       34  GluAlaAlaSerGlyAlaArgValLeuGlyLeuLeuSerLysAlaLeuLeuPro-Th 53
             ||||||||||  |||  | ||||| ::|  ||||||  ||||||  ||||:
Db       16  GAGGCTACAAGTGGCTTGAGAGTGCTGAGAGCGGGCTGCTGAGCCAGGAGCTGTTACCATGG 75

Qy       53  rAlaLeuLeuValThrArgAlaAlaAlaProGlyArgLeuGlnGlyValArgLeuCy 73
             |||||||||| ::| |||||| :::|| ::| | ||||  ||
Db       76  AGCTGCTTGCTAGTGACTACACGCCATGCCGCTACAGGAGGCTGCAGGCAGATTGTG 135

Qy       73  sSerProGlnCysAlaGluValArgGlyPheSerAspLysAspLysLysTyrPheTy 93
             ||||  ||  ::|  |  |||||  |::|    |||  | | ||||  |||
Db      136  CTCGCCACAGTGCGCAGAAATACCGCCGGTTCTCAGACAACAAAAAAAGTATTTCCT 195

Qy       93  rLysPhePheArgAspGluArgArgAlaGluArgAlaTyrArgPheValLysAsnGl 113
             ||||  |  |||||| ||||| |||||||||||  |||  | |||| |||  ||
Db      196  CAAGTTCTTCCGGGACGAGAGGAGGGCAGCAGGCGCTTACCGCTTCGTGAAGAGAACGA 255

CRIGIN
BASE COUNT   190 a   203 c   222 g   182 t    2 others

Alignment Scores:
Pred. No.:                              Length:          798
Score:            9.62e-79              Matches:         196
Percent Similarity: 927.50              Conservative:     25
Best Local Similarity: 82.77%           Mismatches:       43
Query Match:      73.41%                Indels:            5
DB:               28.82%                Gaps:              3
                    9

US-10-029-347-2 (1-625) x AW107005 (1-798)

Qy       32  ProPheGluAlaAlaSerGlyAlaArgValLeuGlyLeuLeuSer-LysAlaLeuLeu 51
             |||||||||| |||  |||  | ||||| ||:  ||||||  |||  :::  ||||||
Db        2  CCTTTGGAGGCTACGAGTGGCTTGAGAGTGCTGAGCGGCGGGGTCTGCTGAGCCAGGAGCTGTTA 61

Qy       52  ProThrAlaLeuLeuValThrArgAlaAlaAlaProGlyArgLeuGlnGlyArg 72
             |||  | | |||||| ||| :::|||||| ::| |  | ||||  |||
Db       62  CT-GTAGCTCGCTTGCTAGTGACTACACGCCCAGGCCGTTACAGGAGGCTGCAGGGCAGA 120

Qy       72  LeuCysSerProGlnCysAlaGluValArgGlyPheSerAspLysAspLysLysTyr 91
             ||||||  ||  ||  | ||||| |||  | ::|  |||  | | ||||  |
Db      121  TTGTGCTCGCCACAGTGCGCAGAAATACCGGCCGGTTCTCCAGACAACAAAAAAGTAT 180

Qy       92  PheTyrLysPhePheArgAspGluArgArgAlaGluArgAlaTyrArgPheValLysGlu 111
             ||||  |  |  |||||| ||||| |||||||||  |||  | |||| |||  |
Db      181  TTCTTCAAGTTCTTCCGGGACGAGAGGAGGGCAGCAGGCGCTTACCGCTTCGTGAAGAG 240

Qy      112  AsnGluThrLeuPheAlaLeuCysPheValProPheValCysTrpIleValCysThrVal 131
             || ||| | ||||| |||  |  ||||| :::| ||  | ||  |||| |||
Db      241  AACGAGACGCTGTTTGCACTGTGCTTTGTGCCC-TCGTGTGCTGGATCGTGTGCACAGTG 300

Qy      132  LeuArgGlnGlnLeuGlyLeuGlyArgAspLeuSerArgThrSerLysThrThrThrSer 151
             |||||| ||||| |||  |||  |||  ||  | ::| | |  || | | | |
Db      301  CTGCAGCAGCTAGAGCTGGGCCGGGATCTCTCCGTACCTNCAAGACCACCACTTCT 360

Qy      152  ValTyrLeuPheIleThrSerValLeuSerSerAlaProValAlaAlaAspGlyProArg 171
             ||||  |||  |  :::| | ||  | ||||  ||  ||  |||
Db      361  GTGTACCTGCTCTTCATCACCAGCAGCTGGAAGTCTGCAGG-- ACCAATGGACCCCCGG 417

Qy      172  LeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGluGluGly--Arg 190
             :| |  |||  :::| |||  | |||||  ||| ::|   ||||
Db      418  GTTCAGGGAGACTGAGAGTGCTAAAAACGCTGAGAGCTACCACCAG 477

Qy      191  ArgAlaGlnPheAlaGlnLysSerGluLeuGluGlyLeuArgGlyValThrGluValGln 210
             |||||| ||| :::| |  | ||||| |::|| :::| ||||  | |  ||||
Db      478  AAAGCACAATTCTCAGAGACAACAAAATTGAAGCTTCAGGTGTCGGTGTCCCAAGTCAG 537

Qy      211  ThrLeuPheLeuSerLysLysGluLeuGluProGlyValLeuGluThrThrTyrGln 230
             |||||| ||||| |::| | ||||||  |  | ::|  |||||  ||::|  |
Db      538  ACAATATTCCTCAACAGAAGAAGATACCAGAGTGCTAAAAACGTGAGCCACCTACCAG 597

Qy      231  PheIleAspGlnSerPheGlnPheLeuGlnLeuLeuAlaLeuSerTyrLeuGluAspGly 250
             ||||| |  ||||  ||  ||||| |||  ::| ||  |  ||
Db      598  TTCATTGACCAGAGCTTCCAGGAG-TCTTGGCTGCATGTCATACCTACTAGAAGCTGAG 657

Qy      251  GlyValProArgThrAlaLeuGlyValGlyThrGluValLeuArgGlyValAspAlaGlnPro 270
             ||||  |  | ||  | ||||| |||  :::
Db      658  CGAACTCCGGGACGCTCG-CAGCGGTGTGCAGAAGCTCCTGAACTACTGACGCGGAGCTA 717

Qy      271  HisSerHisSerLeuValLeuThrArgArgPheGlyLeuPheLeuLeuSerAlaGlnArgMet 290
             ||||  | |  |  ::::||||
Db      718  CGTGGTCATCTTGCACTGACCACCCGATTTCTCTGCCT--GCTAATACACAGAGGGGGCTC 774

modified polylinker; Site 1: Not I; Site 2: Eco RI;
Plasmid DNA from the normalized library NCI_CGAP_Col6 was
prepared, and ss circles were made in vitro. Following HAP
purification, this DNA was used as tracer in a subtractive
hybridization reaction. The driver was PCR-amplified cDNAs
from a pool of 5,000 clones made from the same library
(cloneIDs 105746-1061255, and 114584-145535).
Subtraction by Bento Soares and M. Fatima Bonaldo. "

BASE COUNT      100 a      232 c      245 g      113 t      1 others
ORIGIN

Alignment Scores:
Pred. No.:                                    Length:            691
Score:                       1.93e-69         Matches:           162
Percent Similarity:          830.00           Conservative:        0
Best Local Similarity:       99.39%           Mismatches:          1
Query Match:                 99.39%           Indels:              0
DB:                          25.79%           Gaps:                0
                             9

us-10-029-347-2 (1-625) x AI344276 (1-691)

Qy    1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAlaAspGluLeuProAla    20
Db  197  ...

RESULT 5
BF783992
LOCUS       BF783992      602107835F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:4236272
DEFINITION  938 bp    mRNA    linear    EST 12-JAN-2001
            5', mRNA sequence.
ACCESSION   BF783992
VERSION     BF783992.1  GI:12089028
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 938)
AUTHORS     NIH-MGC http://mgc.nci.nih.gov/.
TITLE       National Institutes of Health, Mammalian Gene Collection (MGC)
JOURNAL     Unpublished

RESULT 4
AI344276
LOCUS       AI344276      691 bp    mRNA    linear    EST 08-APR-1999
DEFINITION  tc03a07.x1 NCI_CGAP_Col6 Homo sapiens cDNA clone IMAGE:2062740 3'
            similar to TR:Q63035 Q63035 VASOPRESSIN RECEPTOR., mRNA sequence.
ACCESSION   AI344276
VERSION     AI344276.1  GI:4081482
KEYWORDS    EST.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 691)
AUTHORS     NCI-CGAP http://www.ncbi.nlm.nih.gov/ncicgap.
TITLE       National Cancer Institute, Cancer Genome Anatomy Project (CGAP),
            Tumor Gene Index
JOURNAL     Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Ilan Kirsch, M.D., Michael R. Emmert-Buck, M.D.,
            Ph.D.
            cDNA Library Preparation: M. Bento Soares, Ph.D.
            cDNA Library Arrayed by: Greg Lennon, Ph.D.
            DNA Sequencing by: Washington University Genome Sequencing Center
            Clone distribution: NCI-CGAP clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            www-bio.llnl.gov/bbrp/image/image.html
            Insert Length: 1242   Std Error: 0.00
            Seq primer: -40UP from Gibco
            High quality sequence stop: 459.
FEATURES             Location/Qualifiers
     source          1..691
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:2062740"
                     /tissue_type="colon tumor, RER+"
                     /lab_host="DH10B"
                     /clone_lib="NCI_CGAP_Col6"
                     /note="Organ: colon; Vector: pT7T3D-Pac (Pharmacia) with a

COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Incyte Genomics, Inc.
             Clone distribution: MGC clone distribution information can be
             found through the I.M.A.G.E. Consortium/LLNL at:
             http://image.llnl.gov
             Plate: LLAM9844 row: e    column: 09
             High quality sequence stop: 660.
FEATURES             Location/Qualifiers
     source
             1..938
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:4236272"
             /lab_host="DH10B (T1 phage-resistant)"
             /clone_lib="NCI_CGAP_Kid14"
             /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
             Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Average insert size 1.75 kb. Constructed by Life
             Technologies. Note: this is a NCI_CGAP Library. |"

BASE COUNT       262 a     232 c     274 g     170 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                3.76e-67          Length:          938
                      808.00            Matches:         195
Percent Similarity:   71.70%            Conservative:    33
Best Local Similarity: 61.32%           Mismatches:      48
Query Match:          25.11%            Indels:          46
DB:                                     Gaps:            6

US-10-029-347-2 (1-625) x BF783992 (1-938)

Qy     50 LeuLeuProThrAlaLeuLeuValThrThrArgAlaAlaAlaProGlyArgLeuGln     69
           |||||||||||||||||||  |||||||||||||||| ||||||||||||||||
Db      3 CTGTTACCAGGAGCTCGCTTGCTCAGTGACTACACGCCATGCCCTACAGGGAGGCTGCAG   62

Qy     70 GlyArgLeuCysSerProG_nCysAlaGluValArgGlyPheSerAspLysAspLysLys   89
           ||||||    ||| |||  ||| :: ||| ||| :::| ||||
Db     63 GGCAGATG-TGCTCGCCACAGTGCGCAGAAA-ACGCGCGCTTCTCAGACAAAGACAAAAA  121

Qy     90 LysTyrPheTyrLysPheArgAspGluArgArgAlaGluArgAlaTyrArgArgPheVal  109
           ||||||  |||| |||  |||:: |||||||||| |||::|| ||||||  |||||
Db    122 AAGTATTTCTTCAAGTTCTTCCGGGACGAGAGGGAAGGCAGAGCGCGCTACCGCTTCGTG  181

Qy    110 LysGluAsnGluThrGluPheAlaLeuCysPheValProPheValCysTrpIleValCys  129
           ||||| :: || ||||||||||| ||| ||| ||| ||| ||| |||
Db    182 AAAGAGAACGAGACGCTGTT-GCACTGTGCTT-GTGCCCTTCGTGTGCTGGATCGTGTGC  239

Qy    130 ThrValLeuArgGlnGluLeuGluGlyArgArgAspLeuSerArgThrSerThrThr    149
           |||::|| ||| |||||||| :: ||| |||| ::: ||| |||||||||||
Db    240 ACAGTGCTGCAGCAGCACTAGAGCTGGGCCGGGATCTTCTCTCGTACCTCCAAGACCACT  299

Qy    150 ThrSerValTyrLeuLeuPheIleThrSerValLeuSerSerAlaProValAlaAspGly  169
           |||||| ||| |||||| ||| ::: ||| ||| ::: :::
Db    300 ACTTCTGTGTACCTGCTCTTCATCACCAGCATGCTGAA-GCTTCTGCAGGCACCAATGGA  358

Qy    170 ProArgLeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGluGlyGlyValLeuGly  189
           ::: ||| |||||| ||| |||::| ||| |||||| ||| |||||||
Db    359 CCCCGGGTTCAGGGAGACCTGAGAACGGTGTGTCGCCTGGCCAGGGAGGGCATCCTGGAT  418

Qy    190 ---ArgArgAlaGlnPheAlaGluLysAspLeuArgAsnLeuCysArgLeuAlaArgGluGlySerLys  209
           :::||:: ||||| ||| ::: ||| ::: |||||||||
Db    419 CACCATAAAGCACAATTCTCAGAAGAGGACCTGGAGAAATTGAAGCTTCGGGGTTCCCAA  478

Qy    209 ValGlnThrLeuPheLeuSerLysLysGluGluProGlyValLeu-GluThrGluValTh  228
           |||||| |||| |||||||||||| :: ||| ||| |||||| :::
Db    479 GTTCAGACAATATTTTCTCAACAGAAAGAGATACCAGGAGTGCTAAAAAACTGAGGTCAC  538

Qy    228 rTyrGlnPheIleAspGlnSerPheGlnGluPheLeuAlaAlaAlaLeuSerTyrLeuLeuGl  248
           :::||:: ||| ||| |||||| ||| |||||||||||||||||
Db    539 CTACCAGTTCATTGACCAGAGTTCCAGGAGTTCTTGGCTGCATTGTCATACCTACTAGA   598

Qy    248 uAsp-GlyGlyValProArgThrAlaAlaGlyValGlyThrLeuArgGlyAlaAspA  268
           :::| ||| |||||||| |||||||||||  |||||||
Db    599 AGCTGAGCGGAACTCCCGGGGAGGCCCCGCCAGGCGGTGTGCAGAAGCTCCGGAACTCGACG  658

Qy    268 la----------------------------GlnProHisSerHisLeuValLeuThr  278
           ||:
Db    659 CGGAACTACGGTGGTTCAATCTTGGCACTGACCACCC--------------------   696

Qy    278 hrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPheG  298
           |||||||| ::: ::::::
Db    697 --GAATTCCTTCTTGGCCTGTAAATACAGGAAGGTTCGGACATGGGAACCC-------   747

Qy    298 lyCysMetValSerGluArgValArgGlnGluAlaLeuArgTrpValGlnGly-----  315
           ---------------------------ATTTGGGTGGGTGGTGGCCACA   769
Db    748

Qy    316 --GlnGlyGlnGlyCysProGly----------ValAlaProGluValThrGluGlyAlaL  332
           |||||: ||||||||:
Db    770 CCTGGGGAAAGGGCCCCGGGGGGGGTCACAGGACGACCCCCCAGGGCACCGGGGGAAAA  829

Qy    332 ysGlyLeuGluAspThrGluProGluGluGlyGluGlyGlu  348
           ::: ::: ::: ::||||||
Db    830 AACGTGGGGGGAATGGGACAAAGAGTGAGGAAAAAGACAAAGGAAGGAAGAGAG  879

RESULT 6
BG971489
LOCUS       BG971489          603 bp     mRNA    linear    EST 12-JUN-2001
DEFINITION  602838184F1 NCI_CGAP_Kid14 Mus musculus cDNA clone IMAGE:497485
            5', mRNA sequence.
            1 (bases 1 to 603)
ACCESSION   BG971489
VERSION     BG971489.1    GI:14359126
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1 (bases 1 to 603)
  AUTHORS   NIH-MGC http://mgc.nci.nih.gov/.
  TITLE     National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL   Unpublished
COMMENT     Contact: Robert Strausberg, Ph.D.
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: Jeffrey E. Green, M.D.
            cDNA Library Preparation: Life Technologies, Inc.
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: Incyte Genomics, Inc.
            Clone distribution: MGC clone distribution information can be
            found through the I.M.A.G.E. Consortium/LLNL at:
            http://image.llnl.gov
            Plate: LLAM10959 row: h   column: 22
            High quality sequence stop: 601.
FEATURES             Location/Qualifiers
     source
             1..603
             /organism="Mus musculus"
             /mol_type="mRNA"
             /strain="FVB/N"
             /db_xref="taxon:10090"
             /clone="IMAGE:497485"
             /lab_host="DH10B (T1 phage-resistant)"
             /clone_lib="NCI_CGAP_Kid14"
             /note="Organ: kidney; Vector: pCMV-SPORT6; Site 1: NotI;
             Site 2: SalI; Cloned unidirectionally. Primer: Oligo dT.
             Average insert size 1.75 kb. Constructed by Life
             Technologies. Note: this is a NCI_CGAP Library. |"
BASE COUNT       152 a     153 c     171 g     127 t
ORIGIN

Alignment Scores:
Pred. No.:              5.89e-66          Length:          603

```
Score:                 793.00        Matches:          160
Percent Similarity:    90.10%        Conservative:      22
Best Local Similarity: 79.21%        Mismatches:        16
Query Match:           24.64%        Indels:             4
DB:                       12         Gaps:               2
```

US-10-029-347-2 (1-625) x BG971489 (1-603)

```
Qy   31  AspProPheGluAlaAla-SerGlyAlaArgValLeuValGlyGlyLeuLeuSerLysAlaIle    50
             ||| |||||||||    |||||||||||||||||||| |||||||||||| |||||||
Db    3  GACCCTTGGAGGTACAATGTGGCTGAGAGTGCTGAGCGGGGTGCTGAGCCAGGAGCT             62

Qy   50  uLeuProThrAlaLeuLeuLeuValThrThrArgAlaAlaAlaProGlyValArgLeuIleGl    70
             ||||| ||||||||||| ||| ||||   ||| ||||||||||||||||||||||
Db   63  GTTACCAGGAGCTGCGCTTGCTAGTGACTACACGCCATGCCGCTACAGGAGGAGGCTGCAGGG    122

Qy   70  yArgLeuCysSerProGlnCysAlaGlu-ValArgGlyPheSerAspLysAspLysLysLysV    90
             | ||||| ||  |||  |||    || |||| | ||||| | ||||||||||||||||
Db  123  CAGATTGTGTCTCGCCACAGTGCGCAGACAATACCGGGCTTCTCAGACAAAGACAAAAAAA    182

Qy   90  ysTyrPheTyrLysPhePheArgArgAspGluArgArgAlaGluArgAlaTyrArgPheValL    110
             || ||||| ||||| || ||||| ||| ||||||||||| ||||||||| |||||| ||
Db  183  AGTATTTCTTCAAGTTCTTCCGGGACGAGAGGAGGGCAGAGCGCGCCTACCGGTTCGTGA    242

Qy  110  ysGluAsnGluThrArgLeuPheAlaLeuCysPheValProPheValCysTrpIleValCysT    130
             |||||||||||| || ||  | ||| |||||  | ||||||||||||||||||||||
Db  243  AAGAGAACGAGACGCGTTTGCACTGTGCTTTGTGCCCTTCGTGTGCTGGATCGTGTGTACGA    302

Qy  130  hrValLeuArgGlnGlnLeuGluGluGlyGlySerArgThrArgLysThrThrThrThrThrT    150
             || |||| |||| ||| ||||  |||  ||   | || |||||   |||   ||
Db  303  CAGTGCTGCGCAGCAGCTAGAGTGGGGCCGGATCCTCTCGTACCTCCAAGACCACTA    362

Qy  150  hrSerValTyrLeuLeuPheIleThrSerValLeuSerSerValAlaProValAlaAspGlyP    170
             ||||| || |||||||||| | ||  ||||||  ||     |||||
Db  363  CTTCTGTGACCTGCTCTTCATCACCAGCAGTGCTGAAGTCTGCAGGC--ACCAATGGAC    419

Qy  170  roArgLeuGlnGlyAspLeuArgLeuLysCysArgLeuAlaArgGluLysArgLeuValLeuGly-    189
             | |||| | ||||| ||||| | || ||||| |||   |||| | ||||  |||||
Db  420  CCCGGTTCAGGGAGGAGCTGAGAACGCTGCGCCTGCTCCAGGAGGGCATCGTGATC    479

Qy  190  --ArgArgAlaGlnPheAlaAlaGluLysLysGlnLeuGluGlnLeuLeuArgGlySerLysV    209
               | ||||| |||||  |    | |||| |||   ||||| |||  ||||  ||||||||
Db  480  ATCATAAAGCACAATTCTCAGAAGAGGACCTGGACAATTGAAGCTTCGGGGTTCCCAAG    539

Qy  209  alGlnThrLeuPheLeuSerLysLysLysLeuProGlyValValValLeuGluThrGluValThr    229
             | ||| ||||| ||| ||||| | || ||  |||||| ||| | ||| |||||||
Db  540  TTCAGACAATATTTCTCAACAAGAAGAGATACCAGGAGTGCTAAAAACTGAGTCACCT    599

Qy  229  yr    229
         ||
Db  600  AC    601
```

```
DNA Sequencing by: Incyte Genomics, Inc.
Clone distribution: MGC clone distribution information can be
found through the I.M.A.G.E. Consortium/LLNL at:
http://image.llnl.gov
Plate: LLAM10990 row: j column: 14
High quality sequence stop: 676.
```

FEATURES             Location/Qualifiers
  source             1..832
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4984429"
                     /lab_host="DH10B (T1 phage-resistant)"
                     /clone_lib="NCI_CGAP_Co24"
                     /note="Organ: colon; Vector: pCMV-SPORT6; Site 1: NotI;
                     Site 2: SalI; Cloned unidirectionally.  Primer: Oligo dT.
                     Average insert size 1.6 kb. Constructed by Life
                     Technologies. Note: this is a NCI_CGAP Library."

BASE COUNT        225 a       211 c       243 g       153 t
ORIGIN

```
Alignment Scores:
Pred. No.:                           Length:            832
Score:                 7.42e-60      Matches:           174
Percent Similarity:    731.50        Conservative:       37
Best Local Similarity: 74.30%        Mismatches:         59
Query Match:           61.27%        Indels:             15
DB:                    22.73%        Gaps:                5
                         12
```

US-10-029-347-2 (1-625).x BG965959 (1-832)

```
Qy  146  SerLysThrThrThrSerValTyrLeuLeuPheIleThrSerValLeuSerSerAlaPro    165
             ||||||||||||||| |||||||| ||| | |||||| ||||  ||
Db    1  TCCAAGACCACTACTTCTGTGACCTGCTCTTCATCACCAGCAGTGCTGAAGTCTGCAGGC     60

Qy  166  ValAlaAspGlyProArgLeuGlnGlyAspLeuArgLeuLysCysArgLeuAlaArgGlu    185
             :|||||||||| |||||||||||:||||||:||||||:|:||||||||||||||
Db   61  --ACCAATGGACCCCGGTTCAGGGAGGAGCTGAGAACGCTGCGCCTGCTGCCAGGAG    117

Qy  186  GlyValLeuGly---ArgArgAlaGlnPheAlaAlaGluLysLysGlnLeuGluGlnLeuLeu    204
             :||| ||||          |||||| |||||| ||||||||||||||||| ||:|||
Db  118  GGCATCGTGATCATCATAAAGCACAATTCTCAGAAGAGGACCTGGACAAATTGAAGCTT    177

Qy  205  ArgGlySerLysValGlnThrLeuPheLeuSerLysLysLysLeuProGlyValValLeuGlu    224
             |||||||||||| |||| |||||||||||||||||||| ||||| |||||||||||
Db  178  CGGGGTTCCCAAGTTCAGACAATATTTCTCAACAAGAAGAGATACCAGGAGTGCTAAAA    237

Qy  225  ThrGluValThrTyrGlnSerPheGlnSerPheGlnSerPheGlyAlaAlaGlyThrLeuLeu    244
             |||||||||||||||  :|||| |||||||||||| :||| || |||||||||||
Db  238  ACTGAGGTCACCTACCAGTTCATTGACCAGAGTTCTTGCTGCATTGTCA    297

Qy  245  TyrLeuLeuGluAspGlyGlyValProArgThrAlaAlaGlyValGlyValGlyThrLeuLeu    264
             ||||||||||| | || | | ||||| ||| |||| | |||||||||||||||
Db  298  TACCTACTAGAAGCTGAGCGAACTCCGGGAGACGCCGTGTGCAGGCCGGTGTGCAGAAGCTCCTG    357

Qy  265  ArgGlyAspAlaGlnProHisSerHisHisValLeuThrThrArgPheLeuPheGlyLeu    284
             |||||||||||||||||| |||| :|||| ||||||||||||  |||||||||
Db  358  AACTCTGACGCGGAGCTACGTGCTCATCTGCACTGACCACCCGATTCCTCTTTGGCCTG    417

Qy  285  LeuSerAlaGluArgMetArgAspIleGluArgHisPheGlyCysMetValSerGluArg    304
             || ::|| ||| | ::|||||| | ||:|||||| ||| |||||:|:|||
Db  418  CTAAATACAGAGGGGCTTCGTGACATTGAAAACCATTTGGCTGTGTGTGGCCAGATCAT    477

Qy  305  ValLysGln-GluAlaLeuArgTrpValGlnGlyGlnGly-----GlnGlyCysProGl    322
             |||||||| :::  ||| || |||||| ::||||||  |||    :|::| |||
Db  478  GTGAAACAAGGACACCTGCGGGTGGGTACAAGGACAGAGCCACCCAAGGGGCCACCAGT    537

Qy  322  yValAlaProGluGlyValThrGluGlyValAlaLysGlyLeuGluAspThrGluProGluGl    342
             :|||| |||||||| | ||||||| | ::|||| |||||||  |||||||
Db  538  AGGGGCAAAAAAGACTGCTGAG------------CTGGAGGACATCGAGGAGCCCAGAGGA    585

Qy  342  uGluGluGluGluPro--AsnTyrProLeuLeuLeuTyrCysLeuTyrGl    361
```

RESULT 7
LOCUS        BG965959               832 bp    mRNA      linear    EST 12-JUN-2001
DEFINITION   602829753F1 NCI_CGAP_Co24 Mus musculus cDNA clone IMAGE:4984429 5',
             mRNA sequence.
ACCESSION    BG965959
VERSION      BG965959.1  GI:14353596
KEYWORDS     EST.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 832)
  AUTHORS    Strausberg,R.
  TITLE      NIH-MGC http://mgc.nci.nih.gov/,
             National Institutes of Health, Mammalian Gene Collection (MGC)
  JOURNAL    Unpublished
COMMENT      Contact: Robert Strausberg, Ph.D.
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

```
Db  586  GGAGGAGGAGGAGGAGGACCACCTCAACTCCGGACTGGAGCTGCTGTACCGCTGTATGA  645
         ||||||| |||||  ||||||                     |||| |||||||||||||
Qy  361  uThrGlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGl  38:

Db  646  GACACAGAGGAGGACGG-ATCAGCGACAGGCTCTCAGCAGCCTCCAGAGATAGTACTGGA  734
         ||||||||||||||||  |||||||||||||||||||||||||||||||||||||||||
Qy  381  nArgVal-ArgPheCysArgMetAsp-ValAlaValLeuSerTyrCysValArgCysCys  400

Db  705  GCGAGTTCAGTTGAGCCGCATGGAACCAGAGACGTCTGAATTACGGCATGCAGCGATGC  764
         |||| |||  |||||  ||||||||||||||||||||||||||||||||||||||||||
Qy  401  ProAla-GlyGlnAla-LeuLeuArgLeuIleSerCysArgLeuValAlaAlaAlaGlyGluGlnLysL  420

Db  765  CCCAGATGGCCAGGCTCCGACCACTGCTGACCTGCGGGCAGATGGCGGCAAAGGAAAGAA  824
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Qy  420  ys  420

Db  825  AA  826
```

RESULT 8
AF054176

LOCUS        AF054176    2218 bp    mRNA    linear    HTC 22-MAY-2001
DEFINITION   Homo sapiens angiotensin/vasopressin receptor AII/AVP mRNA, complete cds.
ACCESSION    AF054176
VERSION      AF054176.1  GI:3341995
KEYWORDS     HTC.
SOURCE       Homo sapiens (human)
ORGANISM     Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2218)
AUTHORS      Mao,M., Fu,G., Wu,J.-S., Zhang,Q.-H., Zhou,J., Kan,L.-X.,
             Huang,Q.-H., He,K.-L., Gu,B.-W., Han,Z.-G., Shen,Y., Gu,J.,
             Yu,Y.-P., Xu,S.-H., Wang,Y., Chen,S.-J. and Chen,Z.
TITLE        Identification of genes expressed in human CD34(+) hematopoietic
             stem/progenitor cells by expressed sequence tags and efficient
             full-length cDNA cloning
JOURNAL      Proc. Natl. Acad. Sci. U.S.A. 95 (14), 8175-8180 (1998)
MEDLINE      98318631
PUBMED       9653160
REFERENCE    2  (bases 1 to 2218)
AUTHORS      Zhang,Q.H., Ye,M., Wu,X.Y., Ren,S.X., Zhao,M., Zhao,C.J., Fu,G.,
             Shen,Y., Fan,H.Y., Lu,G., Zhong,M., Xu,X.R., Han,Z.G., Zhang,J.W.,
             Tao,J., Huang,Q.H., Zhou,J., Hu,G.X., Gu,J., Chen,S.J. and Chen,Z.
TITLE        Cloning and functional analysis of cDNAs with open reading frames
             for 300 previously undefined genes expressed in CD34+ hematopoietic
             stem/progenitor cells
JOURNAL      Genome Res. 10 (10), 1546-1560 (2000)
MEDLINE      20499367
PUBMED       11042152
REFERENCE    3  (bases 1 to 2218)
AUTHORS      Zhang,Q.
TITLE        Direct Submission
JOURNAL      Submitted (18-MAR-1998) Rui-Jin Hospital, Shanghai Second Medical
             University, Shanghai Institute of Hematology, 197, Rui-Jin Road II,
             Shanghai 200025, P. R. China
FEATURES             Location/Qualifiers
     source          1..2218
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /cell_type="CD34+ cell"
     CDS             47..1591
                     /codon_start=1
                     /product="angiotensin/vasopressin receptor AII/AVP"
                     /protein_id="AAC39910.1"
                     /db_xref="GI:3341996"
                     /translation="MCFIPLVCWIVCTGLKQQMESGKSLAQTSKTSTAVVVFFLSS.."
```

(alignment data and remaining sequence text)

```
Qy   375  PheProGluLeuAlaLeuGlnArgValArgPheCysArgMetAspValAlaValLeuSer  394
               ||||||
Db   773  CTCTCCCAGATTGAGATCAAT-----CTCTCCCACCAGAATGGACCACCAGGTTCCTCC  826

Qy   395  TyrCysValArgCysProAlaGlyGlnAlaLeuArgLeuIleSerCysArgLeuVal  414
          :::||
Db   827  TTTGCATTGAGAACTGTCATCGGGTGGAGTCACGTGTCCCTGGGGTTTTCCATAACATG  886

Qy   415  AlaAlaGlnGlnGluLysLysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGly  434
               :::||||||
Db   887  CCCAAGGAGGAGGAGGAGGAGGAGGAAAAGGAAGGCCGACACCTTGAT------  931

Qy   435  GlySerSerGlnGlyThrThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGln  454
                                                          :::
Db   932  -----------------------------------------ATGGTGCAG  940

Qy   455  AlaMetThrAspProLeuCysHisLeuSer----SerLeuThrLeuSerHisCysLysLeu  473
               :::
Db   941  TGTGTCCTCCAAGTCCTCTCATGCTGCCTGTTCTCATGGGTTGGGGCGCTGTGGCCCTC  1000

Qy   474  ProAspAlaValCysArgAspLeuSerGluAla-LeuArgAlaAlaProAlaLeuThrGln  493
               :::||||      |||
Db  1001  TCCCATGAGTGCTGCTTCGACATCTCCTTGGTCCTCAGCAGCAACCAGAGCTGGTGGAG  1060

Qy   494  LeuGlyLeuLeuHisAspArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeu... 513
               |||
Db  1061  CTGGACCTGGAGTGACAACGGCCCTGGGTGACTTCCGAATCAGACTTCTGTGTGTGGGACCTG  1120

Qy   514  AlaTrpProGlnCysArgVal  520
               :::
Db  1121  AAGCACCTGTTGTGCAAATCTGAAGAAGCTCTCGGTTGGCGAATTCTGCCCTACGTCAGTCGT  1180

Qy   521  ----------------GlnThrValArgVal  525
                              :::|||
Db  1181  GTTGTCAGCTTTGTCCTCGGTACTAGCAGCACTAATCAGAATCTCACGCACCTTTACTGCG  1240

Qy   526  GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuVal--GlyMetLeuArgGlnSer  544
               :::      |||
Db  1241  AGGCAACACTCTCGGAGACAAGGGATCAAACTACTCTGTGAGGGACTCTTGCACCCCGAC  1300

Qy   545  ProAlaLeuThrThrLeuAspLeuSerGlyLeuCysGlnLeuProAlaProMet----...  561
               |||      |||
Db  1301  TGCAAGCTTCAGGTGTTGGAATTAGACAACTGCAACCTCAGGTCACACTGCTGCTGGGAT  1360

Qy   561  ...  561
Db  1361  CTTTCCACACTTCTGACCTCCGACCCAGAGCCTGCGGAAAGCTGAGCCTGGCCACAATGAC  1420

Qy   562  --------ValThrTyrLeuCysAlaValLeuGlnHisGlnHisGlyLeuCysGlyLeu  576
                    :::||||||
Db  1421  CTGGGCGACCTGGGGGTCATGAGTTCTGTGAAGTGCTGAAACAGCAGCTGCCTCCTG  1480

Qy   577  GlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerGluLeuGlnLeuGlnAla  596
               |||      :::
Db  1481  CAGAACCTGGGGTGTCTGAAATGTATTTCAATTATGAGACAAAAAGTCGTTAGAAACA  1540

Qy   597  ValLysArgAlaLysProAspLeuValIle  606
               :::||||||
Db  1541  CTTCAAGAGAAAAGCCTGAGCTGACCGTC  1570
```

RESULT 9
BY262618

REFERENCE    1 (bases 1 to 416)
  AUTHORS    Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
             Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
             , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
             Gojobori,T., Baldarelli,R., Hill,D.P., Buit,C., Hume,D.A.,
             Quackenbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
             Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chochia,C., Corbani
             ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
             ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
             , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
             Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
             King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
             ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
             ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
             Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
             Ramachandran,S., Ravasi,T., Reed,J.C., Reed,D.J., Reid,J., Ring
             ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
             ,M., Shimada,K., Sultana,R., Takenaka,Y., Taylor,M.S., Teasdale
             ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
             Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
             ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
             Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
             ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
             , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotani,K., Ishii
             ,Y., Itoh,M., Kagawa,I., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
             ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
             ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
  MEDLINE    22354683
   PUBMED    12466851

COMMENT      Contact: Yoshihide Hayashizaki
             Laboratory for Genome Exploration Research Group, RIKEN Genomic
             Sciences Center(GSC), Yokohama Institute
             The Institute of Physical and Chemical Research (RIKEN)
             1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
             Tel: 81-45-503-9222
             Fax: 81-45-503-9216
             Email: genome-res@gsc.riken.go.jp
             URL:http://genome.gsc.riken.go.jp/
             Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
             ,T., Imotani,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
             , Murata,M., Nakamura,K., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
             , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagani
             ,Y., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
             Submission
             Computational Analysis of Full-Length Mouse cDNAs Compared with
             Human Genome Sequences Mamm. Genome. 12, 673-677 (2001)
             Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new
             genes. Genome Res. 10 (10), 1617-1630 (2000)
             RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer. Genome Res.
             10 (11), 1757-1771 (2000)
             Computer-based methods for the mouse full-length cDNA
             encyclopedia: real-time sequence clustering for construction of a
             nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
             cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in Riken.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Tissues were provided by Michela Fagiolini and Takao K. Hensch (
             Laboratory for Neuronal Circuit Development Brain Science Institute
             RIKEN 2-1 Hirosawa.Wako-shi.Saitama 351-0198 Japan ) whose
             assistance we gratefully acknowledge. Please visit our web site
             (http://genome.gsc.riken.go.jp) for further details.
FEATURES             Location/Qualifiers
     source          1..416
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"

```
                                        /db_xref="taxon:10090"
                                        /clone="K33034H14"
                                        /tissue_type="visual cortex"
                                        /clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT     94 a      109 c     126 g      87 t
ORIGIN

Alignment Scores:
Pred. No.:              9.8e-48          Length:             416
Score:                  601.00          Matches:            119
Percent Similarity:     92.03%          Conservative:         9
Best Local Similarity:  85.51%          Mismatches:          10
Query Match:            18.69%          Indels:               1
DB:                     13              Gaps:                 1

US-10-029-347-2 (1-625) x BY262618 (1-416)

Qy   40  ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThr  59
          ||||||||||||||||| ||||||||||||||:||||||||||||||||||:|||||
Db    3  AGAGTGCTGAGCGGGTGCTGAGCCAGGAGCTGTTACCAGGAGCTCGTTGCTAGTGACT  62

Qy   60  ThrArgAlaAlaAlaProGlyArgLeuGlnGlyAlaArgLeuCysSerProGlnCysAlaGlu  79
          ||:||| |||||| ||||||||||||||||  |||||||||||||||||||
Db   63  ACACGCCATGCCCGCTACAGGGAGGCTGCAGGGCAGATTGCGCTCGGCCACAGTGCGGCAGA  122

Qy   80  ValArgGlyPheSerAspLysAspLysLysLysTyrPheTyrLysPhePheArgAspGlu  99
          :|||:|||||||||||
Db  123  ATACGCGGCTTCTCAGACACAGACAGACAAAAAAGTATTTCTTCAAGTTCTTCCGGAGCGAG  182

Qy  100  ArgArgAlaGlnArgAlaThrArgArgPheValLysGluAsnGluThrGluPheAlaLeuCys  119
          ::|||||||||||||||||||||
Db  183  AGGAAGGCAGAGCCGCGCTTACCGCTTCGTGAAAGAGAACGAGACGCGTGTTTGGACTGTGC  242

Qy  120  PheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGlnLeuGluLeuGly  139
          |||||||||:||||||||||||||||||||||
Db  243  TTTGTGCCTTCGTGTGCTGGATCGTGTGCACAGTGCTGCAGCAGCAGCTAGAGCTGGGC  302

Qy  140  ArgAspLeuSerArgThrSerLysThrThrThrSerValTyrLeuLeuPheIleThrSer  159
          |||||||||||||||||||||||||||
Db  303  CGGGATCTCTCGTACCTCCAAGACCACTACTTCTGTGTACCTGCTTCTCATCACCAGC  362

Qy  160  ValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspAspGlyAspLeuArg  177
          :||||||||
Db  363  ATGCTGAAGTCTCGCAGGC---ACCAATGGACCCCGGGTTCAGGGAGGCTGAGA  413

RESULT 10
BB633515
LOCUS       BB633515                                 670 bp    mRNA    linear    EST 26-OCT-2001
DEFINITION  BB633515 RIKEN full-length enriched, adult male spinal cord Mus
            musculus cDNA clone A330S5K17 5', mRNA sequence.
ACCESSION   BB633515
VERSION     BB633515.1    GI:16469948
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
  ORGANISM  Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 670)
  AUTHORS   Arakawa,T., Carninci,P., Fukuda,S., Furuno,M., Hanagaki,T., Hara,A.
            , Hiramoto,K., Hori,F., Ishii,Y., Ito,M., Kawai,J., Konno,H., Kouda
            ,M., Koya,S., Matsuyama,T., Miyazaki,A., Nomura,K., Ohno,M.,
            Okazaki,Y., Okido,T., Saito,R., Sakai,C., Sakai,K., Sano,H., Sasaki
            ,D., Shibata,K., Shinagawa,A., Shiraki,F., Takahashi,F., Toya,T.,
            Tagami,M., Tagawa,A., Takeda,Y., Tanaka,T., Toya,T.,
            Muramatsu,M. and Hayashizaki,Y.
  TITLE     RIKEN Mouse ESTs (Arakawa,T., et al 2001)
  JOURNAL   Unpublished
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
```

```
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K., Itoh
            ,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. . 10 (10), 1617-1630 (2000)
            wagi,K., Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E.,
            Watahiki,M., Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura
            ,S., Kawai,J., Okazaki,Y., Muramatsu,M. Inoue,Y., Kira,A. and
            Hayashizaki,Y.
            RIKEN integrated sequence analysis (RISA) system--384-format
            sequencing pipeline with 384 multicapillary sequencer. Genome Res.
            10 (12), 1757-1771 (2000)
            Konno,H., Fukunishi,Y., Shibata,K., Itoh,M., Carninci,P., Sugahara
            ,Y. and Hayashizaki,Y.
            Computer-based methods for the mouse full-length cDNA
            encyclopedia: real-time sequence clustering for construction of a
            nonredundant cDNA library. Genome Res. . 11 (2), 281-289 (2001)
            Kondo,S., Shinagawa,A., Saito,T., Kiyosawa,H., Yamanaka,I., Aizawa
            ,K., Fukuda,S., Hara,A., Itoh,M., Kawai,J., Shibata,K. and
            Hayashizaki,Y.
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences. Mamm. Genome. 12, 673-677 (2001)
            Please visit our web site (http://genome.gsc.riken.go.jp) for
            further details.
FEATURES    e mouse tissues.
            Location/Qualifiers
  source    1..670
            /organism="Mus musculus"
            /mol_type="mRNA"
            /db_xref="taxon:10090"
            /clone="A330S5K17"
            /sex="male"
            /tissue_type="spinal cord"
            /dev_stage="adult"
            /lab_host="DH10B"
            /clone_lib="RIKEN full-length enriched, adult male spinal
            cord"
            /note="Site 1: SalI; Site 2: BamHI; cDNA library was
            prepared and sequenced in Mouse Genome Encyclopedia
            Project of Genome Exploration Research Group in Riken
            Genomic Sciences Center and Genome Science Laboratory in
            RIKEN. Division of Experimental Animal Research in Riken
            contributed to prepare mouse tissues. 1st strand cDNA was
            primed with a primer (5'
            GAGAGAGAAGAAGATCCAAGAGCTCTTTTTTTTTTTTTTTTTTVN 3'), cDNA was
            prepared by using trehalose thermo-activated reverse
            transcriptase and subsequently enriched for full-length by
            cap-trapper. cDNA went through one round of normalization
            to Rot = 10.0 and subtraction to Rot = 459.0. Second
            strand cDNA was prepared with the primer adapter of
            sequence (5' GAGAGAGAATTCTGAGTAAT-AAATTATCCCCCCCCCCCC
            3'). cDNA was cleaved with XhoI and BamHI. Vector: a
            modified pBluescript KS(+) after bulk excision from Lambda
            FLC I."
BASE COUNT    143 a      189 c      210 g      128 t
ORIGIN

Alignment Scores:
Pred. No.:              2.03e-47          Length:             670
Score:                  601.00           Matches:            121
Percent Similarity:     86.99%           Conservative:         6
Best Local Similarity:  82.88%           Mismatches:          19
Query Match:            18.68%           Indels:               1
DB:                     10               Gaps:                 0

US-10-029-347-2 (1-625) x BB633515 (1-670)

Qy    1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuLeuAspGlyAlaAspGluLeuProAla  20

Db  233  ATACTGCGCAACCCAATCGCCTACTGTTCATCCTGACGGTGCGACGAGCTGCCACT  292
```

FEATURES          Location/Qualifiers
source            1..3368
                  /organism="Homo sapiens"
                  /mol_type="mRNA"
                  /db_xref="taxon:9606"
                  /clone="IMAGE:4811303"
                  /tissue_type="Placenta, choriocarcinoma"
                  /clone_lib="NIH_MGC_21"
                  /lab_host="DH10B-R"
                  /note="Vector: pOTB7"

BASE COUNT        900 a       803 c       828 g       837 t
ORIGIN

Alignment Scores:
Pred. No.:        1..3368                         Length:          3368
Score:            598.50                          Matches:         193
Percent Similarity:   44.56%                      Conservative:    114
Best Local Similarity:    28.01%                  Mismatches:      251
Query Match:          18.60%                      Indels:          131
DB:               11                              Gaps:            19

US-10-029-347-2 (1-625) x BC021272 (1-3368)

Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheLeuAspGlyAlaAspGlyAlaAspGluLeuPro--- 19
Db    725 ATCGTGTTCACCCGAGGAGACTCTTGTTCGTCATCGACAGCTCGAAGAGCTGCAGGGC 784

Qy     20 AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAla 39
Db    785 GGCTTGAACGACCCGATTCGGATCTGTGTGGTGACTTGATGAGAACGGCCGGTGCAG 844

Qy     40 ArgValLeuGlyGlyLeuLeuSerLysAlaLeuAlaLeuLeuProThrAlaLeuLeuValThr 59
Db    845 GTGCTTCTGAGCAGTTTGCTGAGGAGAAGATGCTCCCGGAGGCCTCCTGCTCATCGCC 904

Qy     60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu 79
Db    905 ATCAAACCCGTGCCCGAAGGAGCTCCGGGATCAGGTGACGATCTGCAGAAATCTACCAG 964

Qy     80 ValArgGlyPheSerAspLysAspLysLysTyrPheTyrLysPheArgAspGlu 99
Db    965 CCCCGGGATTCAACGAGAGTGATAGGTAGGTAGTTTCTGCTGTTTCTCAAGACCCCG 1024

Qy    100 ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys 119
Db   1025 AAAAGAGCCATGGAAGCCTCAATCTTGTAAGAGAAAGTGAACAGCTGTTTCCAATGC 1084

Qy    120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluLeuGly 139
Db   1085 CAAATCCCGCTCCTTCGTGGATCCGTGTACCAGTGTGAACAAGAAGATGCAGAAGGA 1144

Qy    140 ArgAspLeuSerArgThrSerLysThrSerValTyrLeuLeuPheIleThrSer 159
Db   1145 AAAGACCTGGCCCTGACCTGCCAGAGCACTACCTCGTACCTCCTCTTCGTCTTTAAC 1204

Qy    160 ValLeuSerSerAlaProValAlaAspGly------ProArgLeuGlnGlyAspLeuArg 177
Db   1205 ---CTGTTCACACCTGAGGGTGCCGAGGGCCGACTCCGCAAACCCAGCACCAGCTGAAG 1261

Qy    178 AsnLeuCysArgLeuAlaArgGluGlyValLeuGlyArgArgAlaGlnGluLys 197
Db   1262 GCCCGTGCTCCCCTGGCTGCGAGGGGTATGTGGACAGACACATTTGAGTTTTGTGAGAC 1321

Qy    198 GluLeuGluGlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLys 217
Db   1322 GACCTCCGAGAAATGGGTTGTGACGTGACCATCCTGCCGCTG---CTGGGCACCAAG 1378

Qy    218 GluLeuProGlyValLeuGluThrGluValThrTyrGlnPheIleAspSerPheGln 237
Db   1379 ATACTTCTGAAGTACGGGAGCGTGAGCTCCTACGTGTTCCTCCACGTGTGTATCCAG 1438

Qy    238 GluPheLeuAlaLeuSerTyrLeuGluGluAspGly--GlyValProArgThrArgAla 256

Qy     21 LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg 40
Db    293 TTACCATCCTCAGAGGGCCACTCCCTGCAAAGACCCTTTGAGGGCTACAAGTGGCCTGAGA 352

Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuAlaLeuLeuProThrAlaLeuLeuValThrThr 60
Db    353 GTGCTGAGCGGCGGCTGCTGAGCCAGGAGCTGTTACCAGAGAGCTCGCTTGCTAGTGACTACA 412

Qy     61 ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGluVal 80
Db    413 CGCCATGCCGCTACAGGGAGCGCCAGGGCAGATTGTGCTGCCCACAGTGCGCAGAAATA 472

Qy     81 ArgGlyPheSerAspLysAspLysLysTyrPheTyrLysPheArgAspGluArg 100
Db    473 CGCGGCTTCCAGACAAAGACAAAAAAAGTATTTCTACAAGTTCTTCCGGGACGAGAGG 532

Qy    101 ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe 120
Db    533 AAGGCAGAGCGCGCCTACCGTTCGTGAAAGA-ACGAGACGTGTTTGCACTGTGCTTT 591

Qy    121 ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGlnLeuGlyArg 140
Db    592 GTGCCCTTCTGGGTGTGCTGGATCGTGTGCACCAGTGTGCAGCAGCTAGAGCTGGGCCG 65-

Qy    141 AspLeuSerArgThrSer 146
Db    652 GGTCTTCCTCGTACATCC 669

RESULT 11
BC021272

LOCUS       BC021272                3368 bp    mRNA            linear   HTC 18-JAN-2002
DEFINITION  Homo sapiens, clone IMAGE:4811303, mRNA.
ACCESSION   BC021272
VERSION     BC021272.1  GI:18204229
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3368)
  AUTHORS   Strausberg,R.
  TITLE     Direct Submission
  JOURNAL   Submitted (14-JAN-2002) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
            Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
            DNA Sequencing by: National Institutes of Health Intramural
            Sequencing Center (NISC),
            Gaithersburg, Maryland;
            Web site: http://www.nisc.nih.gov/
            Contact: nisc_mgc@nhgri.nih.gov
            Shevchenko,Y., Wetherby,K.D., Beckstrom-Sternberg,S.M.,
            Benjamin,B., Blakesley,R.W., Bouffard,G.G., Brinkley,C., Brooks,S.,
            Dietrich,N.L., Guan,X., Gupta,J., Ho,S.-L., Karlins,E., Legaspi,R.,
            Lim,M., Maduro,Q.L., Masiello,C., Mastrian,S.D., McCloskey,J.C.,
            McDowell,J., Pearson,R., Snyder,B., Stantripop,S., Thomas,P.J.,
            Tiongson,E.E., Touchman,J.W., Tsurgeon,C., Vogt,J.L., Walker,M.A.,
            Zhang,L.-H. and Green,E.D.
COMMENT     Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 39 Row: g Column: 3
            This clone was selected for full length sequencing because it
            passed the following selection criteria: Hexamer frequency ORF
            analysis, GenomeScan gene prediction
            This clone has the following problem: frame shifted.

Qy  540 MetLeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGlyCysG:nLeuProAla 559
        |||||| ::::: :||:: ||||  :||   ||||||  || ::| ||||||||||
Db 2474 ATGCTTCTGCGTAACAGAGACGTGCGCTATCTAGACTCAGTGCCAATGTCCTGAAGGAC 2533

Qy  560 ProMetValThrTyrLeuCysAlaValLeuGlnHisGl:nGlyCysGlyLeuGlnThrLeu 579
        |||||: ||| ||: ::||: :||  ||:|||| ::||:
Db 2534 GAAGGACTGAAAACTCTCGCGAGGCCTTGAAAACTCCGGACTGCTGCCGATTCACTG 2593

Qy  580 -------------------------------------------SerLeuAlaSerValGluLeu 586
                                                     ::|||||||
Db 2594 TGTTTGGTAAAATGTTTATCACTGCTGGCTGCTGTGAGACCTCGCTCGCTCTCATC 2653

Qy  587 SerGluGlnSerLeuGlnLeuGln 595
        ||:::|||:||||||||
Db 2654 AGCAATCAAAACCTGAAGATTCTGCAA 2680

RESULT 22
BC012789
LOCUS       BC012789                              3369 bp    mRNA      linear   HTC 04-MAR-2003
DEFINITION  Homo sapiens, c:one IMAGE:3957172, mRNA.
ACCESSION   BC012789
VERSION     BC012789.1  GI:15215377
KEYWORDS    HTC.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3369)
AUTHORS     Strausberg,R.
TITLE       Direct Submission
JOURNAL     Submitted (15-AUG-2001) National Institutes of Health, Mammalian
            Gene Collection (MGC), Cancer Genomics Office, National Cancer
            Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
            USA
REMARK      NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT     Contact: MGC help desk
            Email: cgapbs-r@mail.nih.gov
            Tissue Procurement: ATCC
            cDNA Library Preparation: Rubin Laboratory
            cDNA Library Arrayed by: The I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            DNA Sequencing by: Institute for Systems Biology
            http://www.systemsbiology.org
            contact: amadan@systemsbiology.org
            Anup Madan, Jessica Fahey, Erin Helton, Mark Ketteman, Anuradha
            Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting
            Clone distribution: MGC clone distribution information can be found
            through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
            Series: IRAL Plate: 25 Row: k Column: 23
            This clone has the following problem: retained intron.
FEATURES             Location/Qualifiers
     source          1..3369
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /clone="IMAGE:3957172"
                     /tissue_type="placenta, choriocarcinoma"
                     /clone_lib="NIH_MGC_21"
                     /lab_host="DH10B-R"
                     /note="Vector: pOTB7"
BASE COUNT      901 a    803 c    828 g    837 t
ORIGIN

Alignment Scores:
Pred. No.:                                    3369
Score:                                        2.88e-46
Percent Similarity:                           598.50
Best Local Similarity:                        44.05%
Query Match:                                  28.01%
DB:                                           18.60%

Length:              3369
Matches:             193
Conservative:        114
Mismatches:          251
Indels:              131
Gaps:                19

US-10-029-347-2 (1-625) x BC012789 (1-3369)

Db 1439 GAGTTCTGTGCGCCTTGTCTATTGCTCAGAGCCACCTGATCATCCTCACCCAGCT 1498
Qy  257 AlaGlyGlyValGlyThrLeuLeuArgGlyAlaGlyAlaGlnAlaGlnPro------HisSerHisLeu 274
        ||||||
Db 1499 GTGGAGATGTGTACAGGAATTGCTAGTTGCCAATTTTGAAAAAGCAAGGAGGACAGACACATTGG 1558
Qy  275 ValLeuThrArgPheLeuPheGlyLeuLeuSerArgAlaGluAlaArgMetArgAspIleGlu 294
        |||||| ||||
Db 1559 ATTTTTTGGGGTGTTTCTACTGGCCTTTTAAATAAAAAAAAGGAACAAGAAAAAAACTGGAT 1618
Qy  295 ArgHisPheGlyCysMetValSerGluArgValLysGlnValAlaLeuArgTrpValGln 314
        ||||||
Db 1619 GCGTTTTTTGGCTTCCAACTGTCCCAAGAGATAAAGCAG------ :657
Qy  315 GlyGlnGlyGlnGlyProGlyValAlaProGluValThrGluGlyAlaLysGlyGlyLeu 334
        |||||
Db 1658 ------------------------------CAAATTCACCAGTGCCTGAAGAGCTTA 1684
Qy  335 GluAspThrGlyGluGluGluProGluGluGluGlyGluGluP:oProLeuGln 354
        ::: ||||||||
Db 1685 GGGGAGCGTGGCAATCCTCAGGGACAGGTGGATTCC------------------TTGGCG 1725
Qy  355 LeuLeuTyrCysLeuTyrGluThrGluGluAspAlaPheValArgGlnAlaLeuCysArg 374
        ||||||
Db 1727 ACATTTTTACTGTCTTTGAAATGCAGGATCCTGCCTTGTGAAGCAGGCAGTG------ 1780
Qy  375 PheProGluLeuAlaLeuGluGlnAlaArgPhe------CysArgMetAspValAla 391
        :|:||
Db :781 ------AACCTCCTCCAGAAGCAACTTCATATTATTGACAACGTGGACTTGGTG 1833
Qy  392 ValLeuSerTyrCysValArgCysValAlaGlyProAlaGlyGlnAlaGlyLeuIleSerSerCys 411
        |||||||||||||||| :::
Db 1832 GTTTCTGCCTACTGCTTAAAATACTGCCCCAGCTTGAGGAAACTC----TGTTTTTCCGTT 1888
Qy  412 ArgLeuValAlaAlaGlnGl-uLysLysLysSerLeuGlyLys------------------------ 426
        ::: ||||||
Db :889 CAAAATGTCTTTTAAGAAAAGAGGATGAACACAGCTCTACGTCGGATTACAGCCTCATCTGT 1948
Qy  427 ------------------------------AlaSerLeuGlyGlySer-------- 436
        ||||
Db :949 TGGCATCACATCTGCTCTGTGGTCACCACCAGCGGGCCACCTCAGAGAGCTCCAGGTGCAG 2008
Qy  426 ------------------------------ 426
Db 2009 GACAGGCACCTCAGCGAGTCGACTTGGACCTGTGTCGTGTAACCAGCTGAGGCATCCCAGC 2068
Qy  427 ---ArgLeuGln-------------------AlaSerLeuGlyGlySer-------- 436
        :|:
Db 2069 TGTCGGCCTTCAGAGAGCTTGGAATAAATAACGTTTCC"TTCGGCCAGAGTGTTC"GCTC 2128
Qy  437 ------------------SerG:nGlyThrThrLysGlnHisLeuProAlaSerLeu------ 448
        ::: |||
Db 2129 TTTGAGGTGCTTTTATCAGGCCAGAATACCTGAGCTTCACCCTCACGAAACTC 2188
Qy  449 ------------LeuHisProLeuPheGlnAlaMetThrArgProLeuCysHisLeuSer 464
        :|:|
Db 2189 TCTCCGTGATGACTTCAGTTCCCTTGTGATGCCTTGAACTACCCAGCAGGCAACGTTAAA 2248
Qy  465 SerLeuHisSerHisCysArgArgAspAlaValCysArgAspLeuSerGluAla 484
        |||||||||||
Db 2249 GAGCTAGCCCTGGTAAATTGTCACCTCTCACCCATTGATTGTGAAGTCCTTGCTGGCCTT 2308
Qy  485 LeuArgAlaAlaProAlaLeuThrGlu----LeuHisAsnArgLeuSerGluAla 504
        |||||| :::
Db 2309 CTAACCAACAACAAGAAGCTGACGTATCTGACACCTGCAAGATAGTCTTT------GACACA 2365
Qy  505 GlyLeuArgMetLeuSerGluGlyLeuGluAlaTrpProGlnCysArgValGlnThrValArg 524
        ::: ||| ||||
Db 2366 GGCCTGCCCTTTTGTGAGGAGCCCTG------------TGCACCCGAGACACGGTCCTG 2413
Qy  525 ValGlnLeu----ProAspProGlnArgGlyLeuGlnTyrLeuValAlaGly 539
        |||||
Db 2414 GTATACCTGCGAGTGTGGCCTTCTGCCACCTCAGCGAGCGAGCAGCGTGCTGCGAATACATCTCTGAA 2473

```
Qy    355 LeuLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheValArgGlnAlaLeuCysArg 374
          ::::::::::|||:::|:::||||:::::
Db   1727 ATATTTACTGCTCTTTGAAATGCAGGATCCTGCCTTTGTGAAGCAGGCAGTG------- 1780

Qy    375 PheProGluLeuAlaLeuGlnArgValArgPhe---------CysArgMetAspValAla 391
          |:::||
Db   1781 ------AACCTCCTCCAAGAGACTAACTTCATATTATTGACAACGTGACTTGGTG      1831

Qy    392 ValLeuSerTyrCysValArgCysCysProLeuAlaGlyAlaGlnAlaLeuArgLeuIleSerCys 411
          ::::||| |||
Db   1832 GTTTCTGCCTACTGCTTAAAATACTCTCCAGCTTGAGGAAACTC---TGT-TTCCGTT   1888

Qy    412 ArgLeuValAlaAlaGlnGluLysLysLysLysSerLeuGlyLys--------------  426
          ::: ||| :::
Db   1889 CAAAATGTCTTTAAGAAGAGGATGAACACAGCTCTACGTCGGATCAGCCTCATCTGT  1948

Qy    426 ------------------------------------------------------------  426
Db   1949 TGGCATCACATCTGCTCTGTCTCACCACCAGCGGGCACCTCAGAGAGTCCAGGTGCAG   2008

Qy    426 ------------------------------------------------------------  426
Db   2009 GACAGCACCCTCAGCGAGTCGCACCCTTGTGACCTGTAACACGTGAGGCATCCCAGC   2068

Qy    427 --AlaSerLeuGlyGlySer-----------------------------------------  436
          |||||
Db   2069 TGTCGCCTTCAGAAGCTTGGAATAAATAACGTTTCCTT-TCTGGCCAGAGTGTTCTGCTC   2128

Qy    437 --------SerGlnGlyThrLysGlnLeuProAlaSerLeu------------------  448
          :::|||
Db   2129 TTTGAGGTGCTCTTTATCAGCCAGACTTGAAATACCTGAGGTTCACCCTCACGAAACTC   2188

Qy    449 --------------LeuHisProLeuPheGlnAlaMetThrAspProCysHisLeuSer  464
          :::
Db   2189 TCTCGTGATGCATCAGG-CCCTGTGATGCCTTGAACTACCCAGAGGCAACGTCAAA   2248

Qy    465 SerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAspLeuSerGluAla  484
          |||||
Db   2249 GAGCTAGCGCTGGTAAATTGTCACCTTCTCACCCATTGATTGGAAG-CCTTGGCCTT   2308

Qy    485 LeuArgAlaAlaAlaProAlaAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGluAla  534
          |||||
Db   2309 CTAACCAACAACAAGCTGACGTATCTGAATGTATCCTGCAACCAGTTA---GACACA   2365

Qy    505 GlyLeuArgMetLeuSerGluGluGlyLeuAlaTrpProGlnCysArgValGlnThrValArg  524
          :::
Db   2366 GGCTGTGCCCT-TTTGTGTGAAGCCCTG---------TGCAGCCAGACACGGTCCTG  2413

Qy    525 ValGlnLeu-------------ProAspProGlnArgGlyLeuGlnTyrLeuValGly  539
          ||| |||
Db   2414 GTATACCTGATGTTGGCTTTCTGCCACCTCAGCGAGCAGTGTCGGAATACATCTCTGAA  2473

Qy    540 MetLeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAla  559
          |||||
Db   2474 ATGCTTCTGCGGTAACAAGAGAGGTGCCGTATCTAGACCTCAGTGCCAATGTCCTGAAGGAC  2533

Qy    560 ProMetValThrTyrLeuAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeu  579
          :::
Db   2534 GAAGGACTGAAAACTCTCTGCGGAGGCCTTGAAACATCCGGACTGCCTCGGATTCACTG  2593

Qy    580 ------------------------------------SerLeuAlaSerValGluLeu  586
          :::
Db   2594 TGTTTGTAAAATGTTTTATCACTGCTGCTGCGCTGTGTGAAGACCTGCGCTTCTGCTCATC  2653

Qy    587 SerGluGlnSerLeuGlnGluGln  595
          |||||
Db   2654 AGCAATCAAAACCTGAAGAATTCTGCAA  2680
```

RESULT 23
AK054264
LOCUS       AK054264      3400 bp    mRNA    linear   HTC 05-DEC-2002
DEFINITION  Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
            full-length enriched library, clone:E330007A02 product:Weakly

```
Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheLeuLeuAspGlyAlaAspGluLeuPro--  19
          :::::|||||:|||:::||:::||::
Db    725 ATCGTGTCTCAACCGGAGAGACTCTTGTTCGTCATCGACAGAGTCGAGAGTGCAGGGC  794

Qy     20 AlaLeuGlyGlyProGluLeuAlaProCysThrAspProPheGluAlaAlaSerGlyAla  39
          |||||:::
Db    785 GGCTTGAACGAACCCGATCGGATCTGTGGTGACTTGATGGAGAAACGGCCGGTGCAG   844

Qy     40 ArgValLeuGlyGlyLeuLeuSerGlyLysAlaAlaLeuLeuProThrAlaLeuLeuLeuValThr  59
          :::||||
Db    845 GTGCTTCTGAGCAGTTGCTGAGGAAGAAGATGGTCCCGGAGCCTCCCTGCTCATCGCC   904

Qy     60 ThrArgAlaAlaLeuAlaProGlyValArgLeuGlnArgLeuCysSerProGlnCysAlaGlu  79
          :::
Db    905 ATCAAACCGTGTGCCGGAAGGAGCTCCGGGATCAGGTGACGATCTCAGAAATCTACCAG  964

Qy     80 ValArgGlyPheSerAspLysLysLysTyrPheTyrLysPheArgPheArgAspGlu  99
          :::|||
Db    965 CCCCGGGATTCAACGAGG-GATAGGTAGTGTATTTCGCTGTTCTTCCAAGACCCG  1024

Qy    100 ArgArgAlaGluAlaAlaTyrArgPheValLysGluAsnGlyThrLeuPheAlaLeuCys  119
          |||||
Db   1025 AAAAGACCCATGAGGAGCCCTTCAAATCTTGAAGAGAAGTGAACAGCTGTTTTCCATATGC  1084

Qy    120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnInLeuGluLeuGly  139
          ||||||
Db   1085 CAAATCCCCTCCTGCTGAATCCTGTGTACCAGTCTGAAGCAAGAGATGCAGAAAGGA  1144

Qy    140 ArgAspLeuSerArgThrSerArgLysThrThrSerValTyrLeuLeuPheIleThrSer  159
          :::||||
Db   1145 AAAGACCTGGCCCTGACCTGCCAGAGCACTACCTCGTGTACCTCGTGTACCTTTAAC  1204

Qy    160 ValLeuSerSerAlaProValAlaAspGly------ProArgLeuGlnGlyLeuArg  177
          |||||
Db   1205 --CTGTTCACACCTGAGGGTGCCGGACCTCCGCAAACCCAGCTGGAG  1261

Qy    178 AsnLeuCysArgLeuAlaArgGluGlyLeuAlaArgArgAlaGlnProGluIleAlaGluLys  197
          ||||||
Db   1262 GCCCTGTCTGCTGGCTGCGAGAGGGTATGTGGACAGACACATTTGAGTTTTGTGAAGAC  1322

Qy    198 GluLeuGlnLeuGlnLeuGluArgGlySerLysValGlnThrLeuPheLeuSerLysLys  217
          :::
Db   1322 GACCTCCGGAGAAATGGGGTTGTTGACGCTGACATCCCCGCGTG---CTGGGCACCAAG  1378

Qy    218 GlnLeuProGlyValLeuGlnThrGluValThrThrTyrGlnPheIlePheTyrGlnInSerPheGln  237
          |||||
Db   1379 ATACTTCTGAAGTACGGGACGCGTGCAGAGCTCCACGTGTTCCTCCCACGTGTGTATCCAG  1438

Qy    238 GluPheLeuAlaAlaLeuSerGluTyrLeuLeuGluAspGly---GlyValProArgThrAla  256
          ||||||
Db   1439 GAGTTCTGTGCCGCCCTGTTCTATTTGCTCAAGAGCCACCTTGATCATCCTCACCCAGCT  1498

Qy    257 AlaGlyGlyValGlyThrLeuArgLeuAspGlyAspGlnAlaGlnInPro------HisSerHisLeu  274
          :::
Db   1499 GTGAGATGTGTACAGGAATTGTACAGTTGCCAATTTGAAAAAGCAAGGAGAGGACCATTGG  1558

Qy    275 ValLeuThrThrArgPheGlyLeuPheLeuLeuSerLeuArgMetArgAspIleGlu  294
          :::
Db   1559 ATTTTTTCGGTGTGTTTCAACTGACCCTTTAAATAAAAAGACAAGAAAAACTGAT  1618

Qy    295 ArgHisPheGlyCysMetValSerGluValSerGlnGluGluAlaLeuArgTrpValGln  314
          |||||
Db   1619 GCGTTTTTGGCTTCGTGCCAGCCTGTTCTATTTGCTCAAGAGATAAAGCAG-----  1657

Qy    315 GlyGlnGlyGlnValGlyThrLeuProGluValAlaProGluValThrGluGluAlaLysGlyLeu  334
          |||||
Db   1658 --------CAAATTCACCAGTGCCTGAAGAGCCTA  1684

Qy    335 GluAspThrGluGluProGlnGluGluGluGlyValGluGlyProAsnTyrProLeuGlu  354
          ::: ::: ||||
Db   1685 GGGGAGCCGTGGCAATCCTCAGGGACCAGGTGGATTCC----------TTGGCG  1726
```

```
ACCESSION    AK054264
VERSION      AK054264.1  GI:26344106
KEYWORDS     HTC; CAP trapper.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1
  AUTHORS    Carninci,P. and Hayashizaki,Y.
  TITLE      High-efficiency full-length cDNA cloning
  JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE    99279253
   PUBMED    10349636
REFERENCE    2
  AUTHORS    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muratsu,M. and Hayashizaki,Y.
  TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE    20499374
   PUBMED    11042159
REFERENCE    3
  AUTHORS    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muratsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE      RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE    20530913
   PUBMED    11076861
REFERENCE    4
  AUTHORS    Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
  TITLE      Functional annotation of a full-length mouse cDNA collection
  JOURNAL    Nature 409 (6821), 685-690 (2001)
  MEDLINE    21095660
   PUBMED    11217851
REFERENCE    5
  AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
  REFERENCE  6  (bases 1 to 3400)
  AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Haragaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiraoka,Y., Hirozane,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,J., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,
             Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
```

```
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muratsu,M. and Hayashizaki,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-JUL-2001) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Itoh,M., Konno,H. Okazaki,Y., Muratsu,M. and Hayashizaki,Y.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3400
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:E330007A02"
                     /db_xref="taxon:10090"
                     /clone="E330007A02"
                     /sex="female"
                     /tissue_type="ovary"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days pregnant adult"
     CDS             76..2553
                     /note="unnamed protein product; Weakly similar to PAN1
                     (SCOP|52047, evidence: SCOP)
                     putative"
                     /codon_start=1
                     /protein_id="BAC35710.1"
                     /db_xref="GI:26344107"
                     /translation="MEHFDPLGPNLKDVLRLKDKTELNNFKRTLRSCCSLPDTYKQINK
                     LTMDLANGAQLAEILTDHCPSAWIKRVTVQLIEEINRVDLAELVVKQIEEAVLKVPEE
                     KVSSKPREPSGTLTFPWNFVQGAKRPEDKQEWENFTKRYTAKOLMEWSESKQAQIFSCAFYI
                     ESYKTLLALCNPKIETPFAHAIVLHGPPGSGKTTWAKQLMEWSESKQAQIFSCAFYI
                     SCREVNNTKPCTFAHLLSMDNPSWRDCVIRDLILGKEFLFVVDGFDELTFPAGALIRD
                     LCGDWNTVKPVEVLLGSLLKRKMAPHATLLVTRTQSLHQIFVMMDQPLLVETLGFLE
                     OEKQEFYFQKYFEDEEGEEEDKGEGKALRALKEVRCNADLYOMASLPTACGIFCLCLEL
                     RMKKGEDLSLTCQTYTSMFLNFLCEVFSSETCEDLHNEEFQLFKKICILAANSLLEQ
                     VPILCEEDFLTLKLNLNNLHPMVCRHILFKDSSTHCLSFICLGIQGLAAIIFVQEL
                     GOESKGVSKYSIQNMLSREARLKNPDLSGLLPFVFGLLNETRIQELKTTFGCQISTEV
                     KRKFLECESGBNKPLLLMNMQBILSGLYBSOEEGFVKEAMVLFEDISLHLKTSTD.
                     HASFCLKNSQNLQTMSLKVEKAVFPENVAALESTAKHQRSPDEQRMLTFWTDFCDTFN
                     SNKKLVFLDIHESFLNSSALEILCEKLPSASCCLQKVVLKNISPDDAYEKLCLIFNGY
                     KTISHLILQGGNJDSMHHSLCEVLKNPACNLKFLSLGSCSTRAQKWDDFFFVLKVNQS
                     LIFLDLTDNSLLDKSAKLLCNIWKEPK"
     polyA_signal    3383..3388
                     /note="putative"
     polyA_site      3400
                     /note="putative"
BASE COUNT   1013 a    755 c    767 g    865 t
ORIGIN

Alignment Scores:
 Pred. No.:                                    3400
 Score:                 1.1e-36    Matches:      183
 Percent Similarity:    499.00    Conservative: 124
 Best Local Similarity: 39.82%    Mismatches:   238
 Query Match:           23.74%    Indels:       226
 DB:                    15.51%    Gaps:          23
                          11

US-10-029-347-2 (1-625) x AK054264 (1-3400)

Qy      6 GlnArgLeuLeuPheI-eLeuAspGlyAlaAspGluLeu-----ProAlaLeuGly 23
```

```
              :::      |:|||:::::::||||||   ||:|||:||      |||:||   ||'
Db        835 AAGGAATTCTTATTTGTGGTAGATGGTTTTGATGAGCTGACATTCCCAGCA---GGAGCT 891

Qy         24 ProGluAlaAlaProCysThrAsp---------------ProPheGluAlaAlaSerGly 38
                             |||   |||               |||   ||'
Db        892 CTGATCCGTGACCTTTGTGGTGACTGGAACACAGTGAAGCCTGTAGAGGTC--------- 942

Qy         39 AlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuVal 58
                    :::||||||   ||||||   :::   :::   |||   |||   |||||:|||
Db        943 ------TTACTGGGCAGTTTGCTGAAGAGGAAGATGGCACCTCATGCCACCCTGCTGGTA 996

Qy         59 ThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAla 78
              |||||||||      :::      ::::::         :::      |||
Db        997 ACCACACGGACACAGTCTTTGCACCAAATCTTTGTCATGATGGATCAGCCACTACTAGTA 1056

Qy         79 GluValArgGlyPheSerAspLysAspLysLysLysTyrPheTyrLysPhePheArgAsp 98
              |||      ||||||   :::::::::|||::::::||||:|   |||:::|||   |:|
Db       1057 GAAACTCTGGGCTTCTTGGAGCAAGAAAAGCAGGAGTACTTTCAGAAATACTTTGAAGAT 1116

Qy         99 GluArg---------------------------ArgAlaGluArgAlaTyrArgPheVal 109
              |||                              :::|||   ||||||   :::   |||
Db       1117 GAGGAGGGTGAGGAGGAGGATAAAGGTGAGGGAAAGGCACTTAGAGCTCTCAAGGAAGTG 1176

Qy        110 LysGluAsnGluThrLeuPheAlaLeuCysPheValProPheValCysTrpIleValCys 129
              :::   |||      |||:::   :::      :::|||      |||   |||   |||
Db       1177 AGATGCAATGCTGATCTATACCAAATGGCCTCACTTCCTACTGCATGTGGGATATTCTGC 1236

Qy        130 ThrValLeuArgGlnGlnLeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThr 149
                    |||      :::::::::   |||   |||||||||   |||   :::|||
Db       1237 CTCTGTCTGGAGCTAAGGATGAAGAAGGGGGAAGACCTGTCTCTGACATGTCAGACATAC 1296

Qy        150 ThrSerValTyrLeuLeuPheIleThrSerValLeuSerSerAlaProValAlaAsp--- 168
              |||:||::::::|||   |||:::      |:|   ||||||            |::
Db       1297 ACCTCTATGTTCCTGAACTTCCTGTGCGAAGTCTTCTCATCAGAAACCTGTGAGGACCAT 1356

Qy        169 ---GlyProArgLeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGluGlyVal 187
                             ||          :::   :::|||   ||||||         :::
Db       1357 CTCAATGAGGAATTCCAAATCTTATTCAAGAAAATATGTATCCTGGCTGCCAATAGTCTC 1416

Qy        188 LeuGlyArgArgAlaGlnPheAlaGluLysGluLeuGluGlnLeuGluLeuArgGlySer 207
              |||   :::               |||::::::         |||:::|||      :::
Db       1417 TTGGAACAGGTGCCCATACTCTGTGAAGAAGACTTCTTGACATTAAAATTGAACCTAAAC 1476

Qy        208 LysValGlnThrLeu------------PheLeuSerLysLysGluLeuProGlyValLeu 223
                 :::      :::               |||   ||:
Db       1477 AATTTGCATCCCATGGTGTGCAGGCACATCCTCTTTAAG--------------------- 1515

Qy        224 GluThrGluValThrTyr------GlnPheIleAspGlnSerPheGlnGluPheLeuAla 241
              ::::::      |:|:::         ||||||            |||:::   ||||||
Db       1516 GACAGTTCCAGTACACACTGTCTCTCTTTCATCTGCCTCGGTATCCAGCAACTCCTGGCT 1575

Qy        242 AlaLeuSerTyrLeuLeuGluAspGly------------GlyValProArgThrAlaAla 257
              ||::::   ::::::   |:    |||            ||||||   :::   :::
Db       1576 GCAATAATATTTGTTCAGGAGCTTGGACAGGAAAGCAAGGGTGTCTCTAAATATAGT--- 1632

Qy        258 GlyGlyValGlyThrLeuLeuArgGlyAspAlaGln---ProHisSerHisLeuValLeu 276
                    :::      :::||       :::|||:::      :::      |:|
Db       1633 ------ATACAGAACATGCTTTCAAGGGAAGCAAGATTAAAAAACCCTGACCTGTCTGGG 1686

Qy        277 ThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHis 296
                       |||:::||||||||||||:::      |||:::::::::::::::
Db       1687 CTGCTGCCATTTGTATTTGGCCTCCTGAATGAGACTCGTATTCAAGAGTTGAAGACTACT 1746

Qy        297 PheGlyCysMetValSerGluArgValLysGlnGluAlaLeuArgTrpValGlnGlyGln 316
              |||||||||   :::|||      ||||||::::::   |||
Db       1747 TTTGGCTGCCAAATATCAACTGAGGTCAAGAGAAAATTCCTG------------------ 1788

Qy        317 GlyGlnGlyCysProGlyValAlaProGluValThrGluGlyAlaLysGlyLeuGluAsp 336

Db       1788 ------------------------------------------------------------ 1788

Qy        337 ThrGluGluProGluGluGluGluGluGlyGluGluProAsnTyrProLeu--------- 353
                             |||   |||   ||||||      |||   ||||||
```

```
Db       1789 ---------------GAGTGTGAGTCAGGGGAA------AATAAACCCTTACTATTACTG 1827

Qy        354 ------------GluLeuLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheValArg 369
                          |||:::|||   ||||||||||||:::|:||||:::   ||||||:::
Db       1828 ATGAATATGCAAGAAATCTTGTCCTGTCTCTATGAATCTCAGGAGGAGGGGTTTGTGAAG 1887

Qy        370 GlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArgValArgPheCysArgMetAsp 389
              :::|||:::      |||   :::::::::|||                        |||
Db       1888 GAAGCCATGGTCCTCTTTGAGGACATCTCCTTGCATTTGAAA------ACCAGCACTGAC 1941

Qy        390 ValAlaValLeuSerTyrCysValArgCysCysProAlaGlyGlnAlaLeuArgLeuIle 409
              :::         |||:::|||::::::               |||   |||:::   :::
Db       1942 CTCATTCATGCTTCATTCTGCCTCAAG---------AACTCTCAGAACTTGCAGACAATG 1992

Qy        410 SerCysArgLeu------------------------ValAlaAlaGlnGluLysLysLys 421
              |||   ::::::                        |||||||||   |||
Db       1993 TCCCTGAAAGTAGAAAAGGCGGTCTTTCCAGAGAATGTTGCTGCACTAGAGTCCACAGCT 2052

Qy        422 Lys--------------------------------------------------------- 422
              |||
Db       2053 AAGCATCAAAGGTCCCCGGATGAACAACGCATGTTAACTTTCTGGACAGACTTTTGTGAC 2112

Qy        423 ------SerLeuGlyLysArgLeu------------------GlnAlaSerLeuGlyGly 434
                    :::      |||:::|||                  ::::::   |||
Db       2113 ACATTTAATTCAAATAAGAAACTGGTATTTCTAGACATCCATGAAAGCTTCCTCAACAGC 2172

Qy        435 GlySerSerGlnGlyThrThrLysGlnLeuProAla------------------------ 446
                 :::   :::         ::::::||||||:::
Db       2173 TCTGCACTGGAGATTCTTTGTGAAAAGCTGCCCTCTGCTTCCTGTTGTCTCCAGAAAGTG 2232

Qy        446 ------------------------------------------------------------ 446

Db       2233 GTCCTCAAAAATATTTCCCCAGATGATGCCTACGAGAAATTGTGCCTCATTTTTAATGGT 2292

Qy        447 ---------------------------------------------SerLeuLeuHisPro 451
                                                           |||:::   |||
Db       2293 TATAAAACCATCTCACATCTGATCCTGCAAGGTGGCAACCTGGACAGCATGCATCACTCA 2352

Qy        452 LeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSerLeu--------------- 466
              |||   :::   :::   :::|||   |||:::|||      |||
Db       2353 CTGTGTGAGGTTCTGAAAAATCCAGCCTGCAACCTGAAGTTTCTCAGCCTGGGATCTTGT 2412

Qy        466 ------------------------------------------------------------ 466

Db       2413 TCTACTGCTGCTCAGAAATGGGATGACTTCTTCCCAGTCCTGAAAGTCAATCAATCCCTG 2472

Qy        466 ------------------------------------------------------------ 466

Db       2473 ATATTTCTGGATCTCACAGACAACAGTCTCTTGGACAAAAGTGCTAAGCTCCTATGTAAT 2532

Qy        467 ------------------------------------ThrLeuSerHisCysLysLeuPro 474
                                                  :::|||   :::|||:::|||
Db       2533 ATTTGGAAGGAACCAAAGTGAATACTGCAGAGGGTATCGTTGGAGAACTGTCAGCTCACT 2592

Qy        475 AspAlaValCysArgAspLeuSerGluAlaLeuArgAlaAlaProAlaLeuThrGluLeu 494
              :::|||   |||:::|||||||||      |||      :::      ||||||   |||
Db       2593 GAAGCCTGCTGCAAGGACCTCTCTTCCGTTTTGATGGTCAGCCGAACTCTGACACACCTC 2652

Qy        495 GlyLeuLeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAla 514
                 |||   :::|||:::|||   :::   |||::::::   |||   |||   ::::::
Db       2653 AGCCTGGCCAACAATAAACTTGGGGATAACGGAGTGAAAAACCTGTGTGAAAGCATAAGC 2712

Qy        515 TrpProGlnCysArgValGlnThr---ValArgValGlnLeuProAspProGlnArgGly 533
                    :::|||   :::||||||   |||                           |||
Db       2713 TGCACGGAGTGCAATCTACAGACCCTGGTGCTGTGGTCCTGTAACATCACCAACGCTGGT 2772

Qy        534 LeuGlnTyrLeuValGlyMetLeuArgGlnSerProAlaLeuThrThrLeuAspLeuSer 553
                    ||||||      ||||||:::|||:::   :::|||      |||||||||
Db       2773 TGCCACTATCTCTCAAAGATGCTCAAACAAACACTGAGCCTGAAACACTTGGATCTGGGG 2832

Qy        554 GlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGlnGly 573
                    ::::::                  :::||||||      |||::::::
Db       2833 CTAAATCGCATCGGAACCAAAGGAGCAAAGTTTCTGTGTGAGGCTTTGAAGAACCCCAAG 2892
```

Qy      574   CysGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGlnGlu 593
              ||||  |::|:|| |||       ::   ::  ||  |:::
Db      2893  AGTAAATGAAAGTCTSGGCTGTGTGGATGCTCCATCACCCACTGAACTGCCAAGAC 2952

Qy      594   LeuGlnAlaVallysArgAlaLysProAspLeu 604
              |:::::|    ::|
Db      2953  TTCTCTGAGACTCTTAGAAGCAACAAGAGCCTG 2985

RESULT :4

BY281505

LOCUS       BY281505                  431 bp     mRNA    linear   EST 1-DEC-2002
DEFINITION  BY281505 RIKEN full-length enriched, visual cortex Mus musculus
            cDNA clone K430354M18 5', mRNA sequence.
ACCESSION   BY281505
VERSION     BY281505.1  GI:26471842
KEYWORDS    EST.
SOURCE      Mus musculus (house mouse)
ORGANISM    Mus musculus
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE   1  (bases 1 to 431)
AUTHORS     Okazaki,Y., Furuno,M., Kasukawa,T., Adachi,J., Bono,H., Kondo,S.,
            Nikaido,I., Osato,N., Saito,R., Suzuki,H., Yamanaka,I., Kiyosawa,H.
            , Yagi,K., Tomaru,Y., Hasegawa,Y., Nogami,A., Schonbach,C.,
            Gojobori,T., Baldarelli,R., Hill,D.P., Bult,C., Hume,D.A.,
            Quackerbush,J., Schriml,L.M., Kanapin,A., Matsuda,H., Batalov,S.,
            Beisel,K.W., Blake,J.A., Bradt,D., Brusic,V., Chothia,C., Corbani
            ,L.E., Cousins,S., Dalla,E., Dragani,T.A., Fletcher,C.F., Forrest
            ,A., Frazer,K.S., Gaasterland,T., Gariboldi,M., Gissi,C., Godzik,A.
            , Gough,J., Grimmond,S., Gustincich,S., Hirokawa,N., Jackson,I.J.,
            Jarvis,E.D., Kanai,A., Kawaji,H., Kawasawa,Y., Kedzierski,R.M.,
            King,B.L., Konagaya,A., Kurochkin,I.V., Lee,Y., Lenhard,B., Lyons
            ,P.A., Maglott,D.R., Maltais,L., Marchionni,L., McKenzie,L., Miki
            ,H., Nagashima,T., Numata,K., Okido,T., Pavan,W.J., Pertea,G.,
            Pesole,G., Petrovsky,N., Pillai,R., Pontius,J.U., Qi,D.,
            Ramachandran,S., Ravasi,T., Reed,C.C., Reed,D.J., Reid,J., Ring
            ,B.Z., Ringwald,M., Sandelin,A., Schneider,C., Semple,C.A., Setou
            ,M., Shimada,H., Suitaraa,R., Takenaka,Y., Taylor,M.S., Teasdale
            ,R.D., Tomita,M., Verardo,R., Wagner,L., Wahlestedt,C., Wang,Y.,
            Watanabe,Y., Wells,C., Wilming,L.G., Wynshaw-Boris,A., Yanagisawa
            ,M., Yang,I., Yang,L., Yuan,Z., Zavolan,M., Zhu,Y., Zimmer,A.,
            Carninci,P., Hayatsu,N., Hirozane-Kishikawa,T., Konno,H., Nakamura
            ,M., Sakazume,N., Sato,K., Shiraki,T., Waki,K., Kawai,J., Aizawa,K.
            , Arakawa,T., Fukuda,S., Hara,A., Hashizume,W., Imotari,K., Ishii
            ,Y., Itoh,M., Kagawa,A., Miyazaki,A., Sakai,K., Sasaki,D., Shibata
            ,K., Shinagawa,A., Yasunishi,A., Yoshino,M., Waterston,R., Lander
            ,E.S., Rogers,J., Birney,E. and Hayashizaki,Y.
TITLE       Analysis of the mouse transcriptome based on functional annotation
            of 60,770 full-length cDNAs
JOURNAL     Nature 420, 563-573 (2002)
MEDLINE     22354683
PUBMED      12466851
COMMENT     Contact: Yoshihide Hayashizaki
            Laboratory for Genome Exploration Research Group, RIKEN Genomic
            Sciences Center(GSC), Yokohama Institute
            The Institute of Physical and Chemical Research (RIKEN)
            1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
            Tel: 81-45-503-9222
            Fax: 81-45-503-9216
            Email: genome-res@gsc.riken.go.jp,
            URL:http://genome.gsc.riken.go.jp/
            Aizawa,K., Akimura,T., Arakawa,T., Carninci,P., Fukuda,S., Hirozane
            ,T., Imotari,K., Ishii,Y., Itoh,M., Kawai,J., Konno,H., Miyazaki,A.
            , Murata,M., Nakamura,M., Nomura,K., Numazaki,R., Ohno,M., Sakai,K.
            , Sakazume,N., Sasaki,D., Sato,K., Shibata,K., Shiraki,T., Tagami
            ,M., Waki,K., Watahiki,A., Muramatsu,M. and Hayashizaki,Y. Direct
            Submission
            Computational Analysis of Full-Length Mouse cDNAs Compared with
            Human Genome Sequences Mamm. Genome 12, 673-677 (2001)
            Normalization and subtraction of cap-trapper-selected cDNAs to
            prepare full-length cDNA libraries for rapid discovery of new
            genes. Genome Res. 10 (10), 1617-1630 (2000);

RIKEN integrated sequence analysis (RISA) system--384-format
sequencing pipeline with 384 multicapillary sequencer. Genome Res.
10 (11), 1757-1771 (2000)
Computer-based methods for the mouse full-length cDNA
encyclopedia: real-time sequence clustering for construction of a
nonredundant cDNA library. Genome Res. 11 (2), 281-289 (2001)
cDNA library was prepared and sequenced in Mouse Genome
Encyclopedia Project of Genome Exploration Research Group in Riken
Genomic Sciences Center and Genome Science Laboratory in RIKEN.
Division of Experimental Animal Research in Riken contributed to
prepare mouse tissues
Tissues were provided by Michela Fagiolini and Takao K. Hensch (
Laboratory for Neuronal Circuit Development Brain Science Institute
RIKEN 2-1 Hirosawa,Wako-shi,Saitama 351-0198 Japan ) whose
assistance we gratefully acknowledge. Please visit our web site
(http://genome.gsc.riken.go.jp) for further details.

FEATURES             Location/Qualifiers
     source          1..431
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="taxon:10090"
                     /clone="K430354M18"
                     /tissue_type="visual cortex"
                     /clone_lib="RIKEN full-length enriched, visual cortex"
BASE COUNT           109 a        113 c        114 g        95 t
ORIGIN

Alignment Scores:
Pred. No.:             1.07e-36             Length:              431
Score:                 487.00               Matches:             101
Percent Similarity:    83.33%               Conservative:        19
Best Local Similarity: 70.14%               Mismatches:          22
Query Match:           15.13%               Indels:              2
                       13                   Gaps:                2

US-10-029-347-2 (1-625) x BY281505 (1-431)

Qy  150  ThrSerValTyrLeuGluPheIleThrSerValLeuSerSerAlaProValAlaAspGly 169
         ||||:||||||||||||||||||||||||||||||||||||    :::::
Db  2    ACTTCTGTGTACCTGCTCTTCATCACCAGCATGCTGAAGTCTGCAGGC---ACCAATGGA 58

Qy  170  ProArgLeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGluGlyValLeuGly 189
         ||||||||||||||||||||||||::|::|||||||||||||||||||||||||
Db  59   CCCCGGGTTCAGGGAGAGCTGAGAACGCTGTGTCGCCTGGCTCAGGGAGGGCATCCTGGAT 118

Qy  190  ---ArgArgAlaGlnPheAlaGluLysGluGlnLeuGluLeuArgGlySerLys 208
            :::|||||||||::|||||||||||||||||||||||||||||:||||||
Db  119  CATCATAAAGCACAATTCTCAGAAGAGGAACCTGGAGAAATTGAAGCTTCGGGGTTCCCAA 178

Qy  209  ValGlnThrLeuPheLeuSerLysGluLysProGlyValLeuProGlyLeuGlnValThr 228
         |||||||||||||||||||||||||||||||||||||||||||||||||:||||||||
Db  179  GTTCAGACAATATTCTCAACAAGAAAGAAGATACCAGGAGTACCAGGAGTCGTAAAACTGAGGTCACC 238

Qy  229  TyrGlnPheIleAspGlnSerPheGlnSerPheLeuAlaAlaLeuSerTyrLeuLeuGlu 248
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  239  TACCAGTTCATTGACCAGAGCTTCCAGAGTTCTTGGCTGCCATTGTCATACCTACTAGAA 298

Qy  249  AspGlyGlyValProArgThrAlaAlaGlyValGlyThrLeuLeuArgGlyValAspAla 268
         ||||:||||:|||||||||||||||||||||||||||||||||||||||||||||||
Db  299  GCTGAGCGAACTCCGGGACCGCCCGCAGGCGGTGTGCAGAAGCTCCTGAACTCTGACGCG 358

Qy  269  GlnProHisSerHisLeuValLeuThrArgPheLeuPheGlyLeuLeuSerAlaGlu 288
         :::|   |||||||||||||||||||||||||||||||||||||||:::|||||
Db  359  GAGCTACGTGGTCATCTTGCACTGACCACCCGATTCCTCTTTGGCTGCTAAATACAGAG 418

Qy  289  ArgMetArgAsp 292
         |||||:|||||
Db  419  GGGCTTCGTGAC 430

RESULT :5

AK087843

LOCUS       AK087843               3577 bp     mRNA    linear   HTC 05-DEC-2002

DEFINITION   Mus musculus 2 days pregnant adult female ovary cDNA, RIKEN
             full-length enriched library, clone:E330028A19 product:Similar to
             PAN2 protein, full insert sequence.
ACCESSION    AK087843
VERSION      AK087843.1  GI:26352787
KEYWORDS     HTC; CAP trapper.
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1
  AUTHORS    Carninci,P. and Hayashizaki,Y.
  TITLE      High-efficiency full-length cDNA cloning
  JOURNAL    Meth. Enzymol. 303, 19-44 (1999)
  MEDLINE    99279253
   PUBMED    10349636
REFERENCE    2
  AUTHORS    Carninci,P., Shibata,Y., Hayatsu,N., Sugahara,Y., Shibata,K.,
             Itoh,M., Konno,H., Okazaki,Y., Muramatsu,M. and Hayashizaki,Y.
  TITLE      Normalization and subtraction of cap-trapper-selected cDNAs to
             prepare full-length cDNA libraries for rapid discovery of new genes
  JOURNAL    Genome Res. 10 (10), 1617-1630 (2000)
  MEDLINE    20499374
   PUBMED    11042159
REFERENCE    3
  AUTHORS    Shibata,K., Itoh,M., Aizawa,K., Nagaoka,S., Sasaki,N., Carninci,P.,
             Konno,H., Akiyama,J., Nishi,K., Kitsunai,T., Tashiro,H., Itoh,M.,
             Sumi,N., Ishii,Y., Nakamura,S., Hazama,M., Nishine,T., Harada,A.,
             Yamamoto,R., Matsumoto,H., Sakaguchi,S., Ikegami,T., Kashiwagi,K.,
             Fujiwake,S., Inoue,K., Togawa,Y., Izawa,M., Ohara,E., Watahiki,M.,
             Yoneda,Y., Ishikawa,T., Ozawa,K., Tanaka,T., Matsuura,S., Kawai,J.,
             Okazaki,Y., Muramatsu,M., Inoue,Y., Kira,A. and Hayashizaki,Y.
  TITLE      RIKEN integrated sequence analysis (RISA) system--384-format
             sequencing pipeline with 384 multicapillary sequencer
  JOURNAL    Genome Res. 10 (11), 1757-1771 (2000)
  MEDLINE    20535923
   PUBMED    11076861
REFERENCE    4
  AUTHORS    Kawai,J., Shinagawa,A., Shibata,K., Yoshino,M., Itoh,M., Ishii,Y.,
             Arakawa,T., Hara,A., Fukunishi,Y., Konno,H., Adachi,J., Fukuda,S.,
             Aizawa,K., Izawa,M., Nishi,K., Kiyosawa,H., Kondo,S., Yamanaka,I.,
             Saito,T., Okazaki,Y., Gojobori,T., Bono,H., Kasukawa,T., Saito,R.,
             Kadota,K., Matsuda,H., Ashburner,M., Batalov,S., Casavant,T.,
             Fleischmann,W., Gaasterland,T., Gissi,C., King,B., Kochiwa,H.,
             Kuehl,P., Lewis,S., Matsuo,Y., Nikaido,I., Pesole,G., Quackenbush,J.,
             Quackenbush,J., Schriml,L.M., Staubli,F., Suzuki,R., Tomita,M.,
             Wagner,L., Washio,T., Sakai,K., Okido,T., Furuno,M., Aono,H.,
             Baldarelli,R., Barsh,G., Blake,J., Boffelli,D., Bojunga,N.,
             Carninci,P., de Bonaldo,M.F., Brownstein,M.J., Bult,C.,
             Fletcher,C., Fujita,M., Gariboldi,M., Gustincich,S., Hill,D.,
             Hofmann,M., Hume,D.A., Kamiya,M., Lee,N.H., Lyons,P.,
             Marchionni,L., Mashima,J., Mazzarelli,J., Mombaerts,P., Nordone,P.,
             Ring,B., Ringwald,M., Rodriguez,I., Sakamoto,N., Sasaki,H.,
             Sato,K., Schonbach,C., Seya,T., Shibata,Y., Storch,K.F., Suzuki,H.,
             Toyo-oka,K., Wang,K.H., Weitz,C., Whittaker,C., Wilming,L.,
             Wynshaw-Boris,A., Yoshida,K., Hasegawa,Y., Kawaji,H., Kohtsuki,S.
             and Hayashizaki,Y.
  TITLE      Functional annotation of a full-length mouse cDNA collection
  JOURNAL    Nature 409 (6821), 685-690 (2001)
  MEDLINE    21085660
   PUBMED    11217851
REFERENCE    5
  AUTHORS    The FANTOM Consortium and the RIKEN Genome Exploration Research
             Group Phase I & II Team.
  TITLE      Analysis of the mouse transcriptome based on functional annotation
             of 60,770 full-length cDNAs
  JOURNAL    Nature 420, 563-573 (2002)
REFERENCE    6  (bases 1 to 3577)
  AUTHORS    Adachi,J., Aizawa,K., Akimura,T., Arakawa,T., Bono,H., Carninci,P.,
             Fukuda,S., Furuno,M., Hanagaki,T., Hara,A., Hashizume,W.,
             Hayashida,K., Hayatsu,N., Hiramoto,K., Hiramoto,K., Hiraoka,T.,
             Hori,F., Imotani,K., Ishii,Y., Itoh,M., Kagawa,I., Kasukawa,T.,
             Katoh,H., Kawai,J., Kojima,Y., Kondo,S., Konno,H., Kouda,M.,

TITLE
JOURNAL      Koya,S., Kurihara,C., Matsuyama,T., Miyazaki,A., Murata,M.,
             Nakamura,M., Nishi,K., Nomura,K., Numazaki,R., Ohno,M., Ohsato,N.,
             Okazaki,Y., Saito,R., Saitoh,H., Sakai,C., Sakai,K., Sakazume,N.,
             Sano,H., Sasaki,D., Shibata,K., Shinagawa,A., Shiraki,T.,
             Sogabe,Y., Tagami,M., Tagawa,A., Takahashi,F., Takaku-Akahira,S.,
             Takeda,Y., Tanaka,T., Tomaru,A., Toya,T., Yasunishi,A.,
             Muramatsu,M. and Hayashizaki,Y.
  TITLE      Direct Submission
  JOURNAL    Submitted (16-APR-2002) Yoshihide Hayashizaki, The Institute of
             Physical and Chemical Research (RIKEN), Laboratory for Genome
             Exploration Research Group, RIKEN Genomic Sciences Center (GSC),
             RIKEN Yokohama Institute; 1-7-22 Suehiro-cho, Tsurumi-ku, Yokohama,
             Kanagawa 230-0045, Japan (E-mail:genome-res@gsc.riken.go.jp,
             URL:http://genome.gsc.riken.go.jp/, Tel:81-45-503-9222,
             Fax:81-45-503-9216)
COMMENT      cDNA library was prepared and sequenced in Mouse Genome
             Encyclopedia Project of Genome Exploration Research Group in Riken
             Genomic Sciences Center and Genome Science Laboratory in RIKEN.
             Division of Experimental Animal Research in Riken contributed to
             prepare mouse tissues.
             Please visit our web site for further details.
             URL:http://genome.gsc.riken.go.jp/
             URL:http://fantom.gsc.riken.go.jp/.
FEATURES             Location/Qualifiers
     source          1..3577
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="C57BL/6J"
                     /db_xref="FANTOM_DB:E330028A19"
                     /db_xref="taxon:10090"
                     /clone="E330028A19"
                     /sex="female"
                     /tissue_type="ovary"
                     /clone_lib="RIKEN full-length enriched mouse cDNA library"
                     /dev_stage="2 days pregnant adult"
     CDS             309..3257
                     /note="unnamed protein product; similar to PAN2 protein
                     (SPTR|Q99MW0, evidence: FASTY, 78.9%ID, 100%length,
                     match=2244)
                     putative"
                     /codon_start=1
                     /protein_id="BAC40024.1"
                     /db_xref="GI:26352788"
                     /translation="MASFFSDFGLMWYLEELNKPVKFKEFLKQEVLQLGLKQVGWT
                     EVKKASRQDLASLLLGHYEEKPAWDMTFRFFQKINRKDLIERAKREIDGCPKLYRAHM
                     KTKMTHDSSRAFTISIQNFLKEKFTEDDYDCFENLFQSKGTESKPQVVFLSSGAGVGK
                     TLMLKRLMLAWIESPVFLHKFSYIFYFCCREVKQLKTASLAELISREWPGSAPIEEI
                     LSKPEKLLFIDSLEGMEBCDLFKWESELCDNCTEKOPVNVLSSLLRRKMLPESSLLI
                     SATPESFEKMXENRIEYTHVKIIKGLKERNIKMSFHRLFQDRNRAHEAFSLVRENEQLF
                     TVCQVPVLCWMVATCLKEEIEKGRDPVSIRCRTTSLYTTHIFNLFIPQNAHSPSKKSQ
                     DQLQGLCSLAAEGMWTDTFVFGEEALRRNGIMDSDIPILLDIGMLINIRESEKSYIFL
                     HPSVQEVCAAIFYLLKSHVDHPSQEVKSIEKLMPAFLKKVKVQWIFFGSFIFGLLHS
                     EQKKLEAFFGHQLSQEIKRQYQCLETISGNEELQEQIDGMKLFYCLFEMDDTFLVE
                     AMNCMEQINFVAKDYSDVIVAAHCLGHCFTLKKLSFSTQNVLSGAQEHSYMERLLTAW
                     NHICSVFIISKDIQELRMKDTNLSESAFSVLYNNLKYHNYTLNVLVANNVFFVCEKYL
                     FFELIQNCNLQHLNLSLTILSHSDVKLLCDVLSQAECNIELVVAACSISSDDCKVFA
                     SVLISNKTLKHLNLASNTLDKGIASLCKSLCHPDCILEHLV.ANGLNEKCWDYLSEV
                     LRRNKTLSHLDISSNDLKDEGLKVLCRALTLPDSVLKSLSRHCLITISGCOLAEVL
                     RNNQNLVSLQVSNNKLEDTGVKLLCDAIKHPNCHLEDLGLEACELTGACCEDLASTFT
                     QCKTLWAMNLLKNALDYNGLVVLCEALKQQTCATVVLGLQITDFDTETQAFLVAEQEK
                     NPCLRILSSL"
     polyA_signal    3557..3562
                     /note="putative"
     polyA_site      3577
                     /note="putative"
BASE COUNT      1060 a    728 c    786 g   1003 t
ORIGIN

Alignment Scores:
Pred. No.:                      1 to 3577
Score:                          3.73e-34                Length:              3577
Percent Similarity:             473.00                  Matches:             174
Best Local Similarity:          39.08%                  Conservative:        116
                                23.45%                  Mismatches:          254

Query Match:        14.70%        Indels:        198
DB:                 11            Gaps:          20

US-10-029-347-2 (1-625) x AK087843 (1-3577)

| 333 | Qy | GlyLeuGluAspThrGluGluProGluGluGluGlyGluGluProAsnTyrPro | 352 |
| 1911 | Db | ACTATAAGTGGCAACGAGGAGCTTCAAGAACAGATAGATGGC | 1952 |

| 1 | Qy | MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuPheLeuAspGlyAlaAspGluLeuProAla | 20 |
|  | Db | ATACTATCCAAACCTGAGAAACCTTATTTATCATCGACAGCTAGAAGGATGGAATGT | 1019 |

| 353 | Qy | LeuGluLeuLeuTyrCysLeuTyrGlnGluThrGlnGluAspAlaPheValArgGlnAlaLeu | 372 |
| 1953 | Db | ATGAAGCTGTTTTACTGTGTCTGTTTGAGATGGACGATGACACCTTCTTAGTAGAGCTATG | 2012 |

| 22 | Qy | LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAla | 39 |
|  | Db | GATTTGTTCAAATGGAATGGAGCTGTGTGATAACTGTACAGAGAAGCCAGTGAAT | 1073 |

| 373 | Qy | CysArgPheProGluLeuAlaLeuLeuAsnArgValArgPheCysArg | 388 Met |
| 2013 | Db | ATGGAACAGATTAACTTTGGCTAAGGATTATTCT | 2054 |

| 40 | Qy | ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValThr | 59 |
|  | Db | GTACTGCTGAGCAGTTTGCTCAGGAGGAAGAGATGCTCCCTGAATCCTCCTCATCTCT | 1133 |

| 389 | Qy | AspValAlaValLeuSerTyrCysValArgCysCysProAlaGlyGlnAlaLeuArgLeu | 408 |
| 2055 | Db | GATGTTATGTTGCTGCCCACTGCTTAAAACACTGT | TTTACACTGAAGAAA 2105 |

| 60 | Qy | ThrArgAlaAlaIleAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu | 79 |
|  | Db | GCTACCCCAGAGAGTTTGAAAAAATGGAGAACAGATTGAATACACACATGTGAAAATA | 1193 |

| 409 | Qy | IleSerCys- | Arg 412 |
| 2106 | Db | CTATCCTTTCAACACCAGAATGTCCTGAGTGGAGCTCAAGAACACAGCTATATGGAAAGG | 2165 |

| 80 | Qy | ValArgGlyPheSerAspLysLysLysLysTyrPheTyrLysPheArgAspGly | 99 |
|  | Db | ATAAAGGGACTTAAGGAGAGAGAAATATTAAGATGAGTTCCACAGATTATTCCAAGATAGG | 1259 |

| 413 | Qy | LeuValAlaAla | GlnGlu 418 |
| 2166 | Db | CCGTCACTGCTTGGAATCATATTTGCTCTGTGTTCATAATCAGTAAGGACATTCAGGAA | 2225 |

| 100 | Qy | ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysCys | 119 |
|  | Db | AACAGAGCCCCATGAAGCCTTCAGTCTGGTGAGAAATGAGCAGCTGTTCACTGTATGA | 1319 |

| 419 | Qy | LysLysLysSerLeuGly- | 425 |
| 2226 | Db | CTCCGAATGAAAGACACTAATCTCAGTGAATCAGCCTTTTCAGTCTTGTATAACAATTTG | 2285 |

| 120 | Qy | PheValProPheValCysSerThrProIleValLeuArgGlnGlnLeuGluLeuGly | 139 |
|  | Db | CAGGTCCCTGCTGCTGCTGATGGTGGCTACTTGCTAAAAGAAGAGATAGAGAAGGGA | 1379 |

| 425 | Qy | ------- |  |
| 2286 | Db | AAGTACCACAACTACACCCTTAACGTTCTGTGGCAAATAATGTGTTCTTCGTATGTGAA | 2345 |

| 140 | Qy | ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu- | 154 |
|  | Db | AGAGACCCAGTCCCATTGCCGTGTGACTACCTCCGTATACCACTCACATCTCAAT | 1439 |

| 426 | Qy | LysArgLeuGlnAlaSerLeu | 432 |
| 2346 | Db | AAGTACCTGTTCTTGAGTGATTCAGAACTGAATTACAGACACTTGAACCTCAGCCTC | 2405 |

| 155 | Qy | LeuPheIleThrSerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGly | 174 |
|  | Db | TTGTTCATTCCCAAAATGCCAATTCCCAGTAAGAAGACCAGCCCAGCTGCAGGGC | 1499 |

| 433 | Qy | GlyGlyGlySer- | 436 |
| 2406 | Db | ACAATTCTCTCCCACAGTGATGTGAACTGTTGTGTGATGTGTTGAGCCAGGCGAGAGTGC | 2465 |

| 175 | Qy | AspLeuArgAsnLeuCysArgArgLeuLeuAlaArgGluGlyValValLeuGlyValArgArgAlaGlnPhe | 194 |
|  | Db | TTATGTTCTGTGGCCGCTGAGGGCATGTGACTGTGGACACGTTGTGTTT | 1547 |

| 436 | Qy | ------- |  |
| 2466 | Db | AACATAGAAGAACTTGTGTGTAGCAGCTCGTAGTATTTCATCTGATGACTCCAAGGTGTTT | 2525 |

| 195 | Qy | AlaGluLysGluGluGlnLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeu | 214 |
|  | Db | GGTGAGGAGGCTCCAGGAGAAATGGGATCATGGGATTCTGGATTACATACTA---CTG | 1604 |

| 437 | Qy | SerGlnGlyThrThrLysGlnLeuProAlaSerLeu- | 448 |
| 2526 | Db | GCCTCTGTCCTGATCAGCAACAAGACATCTTAATTTGGCATCCAACACCTTG | 2585 |

| 215 | Qy | SerLysLysGluLeuProGlyValLeuGluThrGluValThrThrTyrGlnPheIleAspGln | 234 |
|  | Db | GACAATGGAATGCTTATAAAATACAGGAGAATCTGAAAAATCTTATATTTTCCATCCA | 1664 |

| 449 | Qy | LeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSer | 465 |
| 2586 | Db | GACAAAGGAATAGCCTCACTGTGCAAGTCTTTGTGCCACCCAGACTGCTGCAGGCAC | 2645 |

| 235 | Qy | SerPheGlnGluPheLeuAlaAlaLeuSerTyrLeuLeuGluLeuAspGly--GlyValPro | 253 |
|  | Db | TTAGTTTTGGCCAACTGCTCCCCTAATGCATCTTTATCTGCTAAAGAGCCACGTGGACCACCCT | 1724 |

| 466 | Qy | LeuThrLeuSerHisCysValCysArgProAspAlaValCysArgAspLeuSerGlyAlaLeu | 495 |
| 2646 | Db | TTGTTTTGGCCAACTGCTCCCCTAATGAGAAATGCTGGGATTACCTTTCAGAAGTTCT | 2705 |

| 254 | Qy | ArgThrAlaAlaGlyValGlyValGlyThrArgLeuArgGlyValAspProAlaGlnPro--HisSer | 272 |
|  | Db | AGCCAGGAGGTTAAAAGTATAGAGAACTCATGTTGCATTTCTAAAGAAGTCAAAGTA | 1784 |

| 486 | Qy | ArgAlaAlaProAlaLeuThrGluLeuLeuGlyLeuLeuHisAsnArgLeuSerGluSerGluAlaGly | 505 |
| 2706 | Db | AGGCGAACAGAGTTCCTGAGCTGAGTTCACTAAGTTGAGACATTGTCTCATCACCATTAGT | 2765 |

| 273 | Qy | HisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerSerAlaGluGluArgMe-ArgAsp | 292 |
|  | Db | GACATTGGAATGCTTATAAAATACAGGGAGAATCTGAAGGTGAAGAGAGAAGCACAAAAAAAG | 1844 |

| 506 | Qy | LeuArgMetLeuSerGluGlyLeuGlyMetLeuValGlnSerProAlaLeuThrValArgVal- | 525 |
| 2766 | Db | GACAAAGGAATAGCCTCACTGTGCAAGTCTTTGTGCCACCCAGACTGCTGCAGGCAC | 2792 |

| 293 | Qy | IleGluArgHisPheGlyCysMetValSerGluArgValLysGluGlnAlaAlaLeuArgTrp | 312 |
|  | Db | AGCCAGGAGGTTAAAAGTATAGAGAACTCATGTTGCATTTCTAAAGAAGTCAAAGTA | 1904 |

| 526 | Qy | GlnLeuProAspProGln- | Arg 532 |
| 2793 | Db | ---CTCCCAGACAGTGTCCTGAGTCACTAAGTTGAGACATTGTCTCATCACCATTAGT | 2849 |

| 313 | Qy | ValGlnGlyGlnGlyGlnGlyCysProGlyValAlaProGluValThrGluGlyValAlaLys | 332 |
|  | Db | CTAGAGGCGGTTTTTGGCCACCAGTGTCCCAAGAAATAAAACGTCAGTTGTATCAGTGC | 1904 |

| 533 | Qy | GlyLeuGlnTyrLeuValGlyMetLeuArgGlnSerProAlaLeuThrArgLeuAspLeu- | 552 |
| 2850 | Db | GGTTGCCAGGACCTGGCTAGAAGTCTTGAGGAACAACCAGGAACCTGGTCAGCCTACAAGTT | 2909 |

| 1905 | Db | CTGGAA------- | 1910 |

| 553 | Qy | SerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGln | 572 |

```
                   :::    :  :          ::::           |||||:            :::::::::|
Db    2910 TGAACAATAAACTAGAGAGACTGGTGTGAAGCTGCTGTGTGATGCTATAAAACATCCC 2969
                |||  |||||  :::    |||   |||    |||||||||:::
Qy     573 GlyCysGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGln 592
               ||| |:::: |||  :::  ||| |||||||:::
Db    2970 AACTGCCACTTAGAGGATCCTGGGTTGGGTTGGAAGCCTGTGAGCTAACTGGTGCCTGTTGTGTGAG 3029

Qy     593 GluLeu 594
              :::|||
Db    3030 GACCTT 3035
```

Search completed: October  3,  2003, 16:42:47
Job time : 3502 secs

OM protein - nucleic search, using frame_plus_p2n model.

Copyright  GenCore version 5.1.6
Copyright (c) 1993 - 2003  Compugen Ltd.

Run on:       October  3, 2003, 11:59:35 ; Search time 6351 Seconds
                                           (without alignments)
                                           4025.904 Million cell updates/sec

Title:        US-10-029-347-2
Perfect score: 3218
Sequence:     1 MLAQPQRLLFILDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table: BLOSUM62
               Xgapop 10.0 , Xgapext 0.5
               Ygapop 10.0 , Ygapext 0.5
               Fgapop  6.0 , Fgapext 7.0
               Delop   6.0 , Delext  7.0

Searched:     288711 seqs, 2045813386 residues

Total number of hits satisfying chosen parameters:         5777422

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:
-MODEL=frame+_p2n.model -DEV=xlh
-Q=/cgn2_i/USPTO.spool/US10029347/runat_02102003_173436_17076/app.query.fasta_1.775
-DB=GenEmbl -QFMT=fastap -SUFFIX=rge -MINMATCH=0.1 -LOOPCL=0 -LOOPEXT=0
-UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi -LIST=45
-DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15 -MODE=LOCAL
-OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10029347 @CGN_1_1_3963 @runat_02102003_173436_17076 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMECUT=120 -WARN_TIMECUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :     GenEmbl:*
    1: gb_ba:*
    2: gb_htg:*
    3: gb_in:*
    4: gb_om:*
    5: gb_ov:*
    6: gb_pat:*
    7: gb_ph:*
    8: gb_pl:*
    9: gb_pr:*
   10: gb_ro:*
   11: gb_sts:*
   12: gb_sy:*
   13: gb_un:*
   14: gb_vi:*
   15: em_ba:*
   16: em_fun:*
   17: em_hum:*
   18: em_in:*
   19: em_mu:*
   20: em_om:*
   21: em_or:*
   22: em_ov:*
   23: em_pat:*
   24: em_ph:*
   25: em_pl:*
   26: em_ro:*
   27: em_sts:*
   28: em_un:*

   29: em_vi:*
   30: em_htg_hum:*
   31: em_htg_inv:*
   32: em_htg_other:*
   33: em_htg_mus:*
   34: em_htg_pln:*
   35: em_htg_rod:*
   36: em_htg_mam:*
   37: em_htg_vrt:*
   38: em_sy:*
   39: em_htgo_hum:*
   40: em_htgo_mus:*
   41: em_htgo_other:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 3218 | 100.0 | 2689 | 6 | AX663294 | AX663294 Sequence |
| 2 | 3218 | 100.0 | 3365 | 6 | AX168123 | AX168123 Sequence |
| 3 | 3207.5 | 99.7 | 2679 | 9 | AF479748 | AF479748 Homo sapi |
| 4 | 3207.5 | 99.7 | 2679 | 9 | AY154461 | AY154461 Homo sapi |
| 5 | 3170 | 98.5 | 2763 | 9 | AX663298 | AX663298 Sequence |
| 6 | 2882 | 89.6 | 5553 | 6 | AX088215 | AX088215 Sequence |
| 7 | 2882 | 89.6 | 5556 | 6 | AX459875 | AX459875 Sequence |
| 8 | 2882 | 89.6 | 6461 | 6 | AX088208 | AX088208 Sequence |
| 9 | 2622.5 | 81.5 | 95038 | 9 | AF015416 | AF015416 Homo sapi |
| 10 | 2619.5 | 81.4 | 143835 | 9 | AC136475 | AC136475 Homo sapi |
| 11 | 2617.5 | 81.3 | 156819 | 2 | AP006283 | AP006283 Homo sapi |
| 12 | 2209 | 68.6 | 2054 | 9 | AX663299 | AX663299 Sequence |
| 13 | 2020 | 62.8 | 3454 | 10 | BC031139 | BC031139 Mus muscu |
| 14 | 1958 | 60.8 | 2354 | 10 | BC013519 | BC013519 Mus muscu |
| 15 | 1746.5 | 54.3 | 238533 | 2 | AC132968 | AC132968 Rattus no |
| 16 | 1737 | 54.0 | 2257 | 10 | RATAIIR | M85183 Rattus norv |
| 17 | 1558.5 | 48.4 | 251206 | 2 | AC107815 | AC107815 Mus muscu |
| 18 | 1120 | 34.8 | 61320 | 2 | AC135992 | AC135992 Homo sapi |
| 19 | 835 | 25.9 | 3913 | 10 | AF486632 | AF486632 Mus muscu |
| 20 | 831 | 25.8 | 3221 | 9 | AY116207 | AY116207 Homo sapi |
| 21 | 826.5 | 25.7 | 3108 | 6 | AX417214 | AX417214 Sequence |
| 22 | 826.5 | 25.7 | 3300 | 6 | AX459869 | AX459869 Sequence |
| 23 | 826.5 | 25.7 | 3466 | 9 | AK095460 | AK095460 Homo sapi |
| 24 | 826.5 | 25.7 | 3563 | 9 | AY116205 | AY116205 Homo sapi |
| 25 | 826.5 | 25.7 | 3731 | 9 | AY116204 | AY116204 Homo sapi |
| 26 | 826 | 25.7 | 3395 | 9 | AY116206 | AY116206 Homo sapi |
| 27 | 825.5 | 25.7 | 3507 | 9 | BC028069 | BC028069 Homo sapi |
| 28 | 825.5 | 25.7 | 3827 | 9 | AY095146 | AY095146 Homo sapi |
| 29 | 825.5 | 25.7 | 3827 | 9 | AY154467 | AY154467 Homo sapi |
| 30 | 818.5 | 25.4 | 3237 | 9 | AY092033 | AY092033 Homo sapi |
| 31 | 814.5 | 25.3 | 3828 | 9 | AF427617 | AF427617 Homo sapi |
| 32 | 804.5 | 25.0 | 3579 | 9 | AF468522 | AF468522 Homo sapi |
| 33 | 800.5 | 24.9 | 3102 | 6 | AX299762 | AX299762 Sequence |
| 34 | 800.5 | 24.9 | 3857 | 9 | AX299760 | AX299760 Sequence |
| 35 | 800.5 | 24.9 | 3857 | 9 | AF420469 | AF420469 Homo sapi |
| 36 | 800.5 | 24.9 | 4170 | 9 | AF410477 | AF410477 Homo sapi |
| 37 | 762 | 23.7 | 4123 | 9 | HSM800983 | HSM800983 Homo sapi |
| 38 | 762 | 23.7 | 4290 | 9 | AF229059 | AF229059 Homo sapi |
| 39 | 762 | 23.7 | 4422 | 6 | AX089761 | AX089761 Sequence |
| 40 | 762 | 23.7 | 4422 | 6 | AF229060 | AF229060 Homo sapi |
| 41 | 762 | 23.7 | 4556 | 6 | AX089769 | AX089769 Sequence |
| 42 | 762 | 23.7 | 5122 | 6 | AX207661 | AX207661 Sequence |
| 43 | 762 | 23.7 | 5306 | 9 | BC051787 | BC051787 Homo sapi |
| 44 | 762 | 23.7 | 5404 | 9 | AF310105 | AF310105 Homo sapi |
| 45 | 762 | 23.7 | 5444 | 9 | AB023143 | AB023143 Homo sapi |

ALIGNMENTS

RESULT 1

```
AX663294
LOCUS       AX663294                2689 bp    DNA     linear   PAT 22-MAR-2003
DEFINITION  Sequence 1 from Patent WO2061086.
ACCESSION   AX663294
VERSION     AX663294.1  GI:23163660
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Feder,J., Ramanathan,C. and Mintier,G.
  TITLE     Human leucine-rich repeat containing protein, expressed
            predominantly in small intestine, HLRSI.
  JOURNAL   Patent: WO 02061086-A 1 08-AUG-2002;
            Bristol-Myers Squibb Company (US)
FEATURES             Location/Qualifiers
     source          1..2689
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
     CDS             75..1952
                     /note="unnamed protein product"
                     /codon_start=1
                     /protein_id="CAD80012.1"
                     /db_xref="GI:29163661"
                     /translation="MLAQPQRLLFILDGADELPALGGPEAAPCDPFEAASGARVLGG
                     LLSKALLPTALLVVTRAAAPGRLGCICSPCCAEVRGFSFDKDKKKYFKFFRRERRA
                     ERAYRFVKENETLFALCFVPFVCKWIVCTVLRQQELEGRDLSRTSKTTTSVYLLFITSV
                     SSAPVADGPRLCGDLRNCRLAREGVLGRRAQFAEKELEQLERGSRVCTLFLSKKE
                     LPGVLETEVTYQFIDQSFQEFLAALSYLLEDGGVPRTAAGGVGTLLRGDAQPHSLVL
                     TTRTLFGLLSAERMRDIERHFGCVVSERVKQEALRWVQGQGQGCPGVAPEVTEGAKGL
                     EDTEPEEEEGEEPNYPLELLYCLYETQEDAFVRQALCRFPELALQRVRFCRMDVAV
                     LSYCVRCCPAGQALRLISCRLVAAQEKKKKSLGKRLQASLGGGSGGTTKQLPASLJH
                     PLFQAMTDPLCHLSSLTLSHCKLPDAVCRDLSEALRAAPALTEGLLHNRLSEAGLRM
                     LSEGLAWPGCRVGTVRVQLPDPQRGLQYLVGMLRQSPALTTLDLSGCQLPAPMYTYLC
                     AVLQRQGCGLQTLSLASVESEQSLQELQAVKRAKPDLVTHPRLDGHPQPPKELIST
                     F"
BASE COUNT      534 a    861 c    844 g    450 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:                    Length:          2689
Score:          9.88e-236     Matches:         625
Score:          3218.00       Conservative:    0
Percent Similarity:  100.00%  Mismatches:      0
Best Local Similarity: 100.00% Indels:         0
Query Match:    100.00%       Gaps:            0
DB:                           6
```

US-10-029-347-2 (1-625): x AX663294  (1-2689)

```
Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla    20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 ATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGGGCGGACGAGCTGCCGGCG   134

Qy     21 LeuGlyGlyProGluAlaAlaProCysAspProPheGluAlaAlaSerGlyAlaArg     40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 CTGGGGGGGCCCGAGGCGGCGCCGTGCGACCCGTTCGAGGCGGCCAGCGGCGCGCG   194

Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValValThrThr    60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 GTGCTAGGCGGCGGCTGCTGAGTAAGGCGCTGCTGCCCACGGCCCTCCTGCTGTGACCACG   254

Qy     61 ArgAlaAlaAlaProGlyArgLeuGlyCysIleCysSerProCysCysAlaGluVal     80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    255 CGCGCCGCCGCCCCCGGGAGGCTGCAGGGCGTGTTCCCCGACCCCGAGTGCGGCCGAGCTG   314

Qy     81 ArgGlyPheSerPheAspLysLysLysLysLysTyrPheLysPhePheArgArgGluArg   100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    315 CGCGGCTTCTCCGACAAGAAGAAGAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGAGG   374

Qy    101 ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe   120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    375 AGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC   434

Qy    121 ValProPheValCysLysTrpIleValCysThrValLeuArgGlnGlnGluLeuGlyArg   140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    435 GTGCCCTTCGTGTGCAAGTGGATCGTGTGCACCGTGCTGCGCCAGCAGGAGCTGGGTCGG   494

Qy    141 AspLeuSerArgThrSerLysThrThrThrSerValTyrLeuLeuPheIleThrSerVal   160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    495 GACCTGTCGCGCACGTCCAAGACCACCACCAGTGTGTACCTGCTTTTCATCACCAGCGTT   554

Qy    161 LeuSerSerAlaProValAlaAspGlyProArgLeuGlnAspGlyArgAsnLeuCys     180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    555 CTGAGCTCGGCTCCGGTAGCCGACGGGCCGCGGCTGCAGGGCGACCTGCGCAATCTGTGC   614

Qy    181 ArgLeuAlaArgGluGlyValLeuGlyArgArgAlaGlnPheAlaGluLysGluLeuGlu   200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    615 CGCCTGGCCCGCGAGGGCGTCCTGGACCGCAGGGCGCAGTTGCCGAGAAGAACTGGAG   674

Qy    201 GlnLeuGluArgGlySerLysValCysGlnThrLeuPheLeuSerLysLysGluLeuPro   220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    675 CAACTGGAGCTTCGTGCTCCAAAGTGCACGCTGTTTCTCAGCAAAAAGGAGCTGCCG   734

Qy    221 GlyValLeuGluThrGluValThrTyrGlnPheIleAspGlnSerPheGlnGluPheLeu   240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    735 GGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGTTCCAGGAGTTCCTC   794

Qy    241 AlaAlaLeuSerTyrLeuLeuGluAspGlyGlyValProArgThrAlaAlaGlyGlyVal   260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    795 GCGGCACTGTCCTACCTGCTGGAGGACGGGGGTGCCCAGGACCGCGGCTGGGCGGCGTT   854

Qy    261 GlyThrLeuLeuArgGlyAspAlaGlnProHisSerLeuValLeuThrThrArgThrPhe   280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    855 GGGACACTCCGGCTGGGGAGCGCCCAGCCCCACAGCCACTTGGTGCTCACCACCCGCTTC   914

Qy    281 LeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPheGlyCysMet   300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    915 CTCTTCGGACTGCTGTCAGCGGAGCGGATGCGGGACATCGAGCGGCACTTCGGGTGCATG   974

Qy    301 ValSerGluArgValLysGlnGluAlaLeuArgTrpValGlnGlyGlnGlyGlnGlyCys   320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    975 GTTTCAGAGCGCGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGACAGGGACAGGGCTGC  1034

Qy    321 ProGlyValAlaProGluValThrGluGlyAlaLysGlyLeuGluAspThrGluGluPro   340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1035 CCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAGGGCTCGAGGACACCGAGGAGCCA  1094

Qy    341 GluGluGluGluGlyGluGluProAsnTyrProLeuGluLeuLeuTyrCysLeuTyr     360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1095 GAGGAGGAGGAGGGAGAGGAGCCCAACTACCCCCTGGAGTTGCTGTACTGCCTGTAC  1154

Qy    361 GluThrGlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu   380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1155 GAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCGCTG  1214

Qy    381 GlnArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysCys   400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1215 CAGCGGGTGCGCTTCTGCCGCATGGACGTGGCGGTGCTGTCCTACTGCGTGCGGTGCTGC  1274

Qy    401 ProAlaGlyGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnGluLysLys   420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1275 CCTGCTGGACAGGCACTGCGGCTCATCAGCTGCAGATTGGTTGGTGCGCAGGAAGAAG  1334

Qy    421 LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGlyGlySerGlyGlyThr     440
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1335 AAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCAGTGGCACC  1394

Qy    441 ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeu   460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1395 ACAAAACAACTGCCAGCCTCCCTTCTCCATCCACTCTTTCAGGCAATGACTGACCCACTG  1454

Qy    461 CysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAsp   480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   1455 TGCCACTCGAGCAGCCTCACCCTGTCCCACTGCAAACTCCCTGACGCGGTCTGCCGAGAC  1514
```

```
Qy    481  LeuArgGluAlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuHisAsnArg  500
Db   1515  CTTTCTGAGGGCCCTGAGGGCAGGCCCCCACTGACGGAGCTGGGCCTCCTCCACACAGG  1574

Qy    501  LeuSerGluAlaGlyLeuLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgVal  520
Db   1575  CTCAGTGAGGGGACTGCGTATGCTGAGTGAGGGCTAGCCTGGCCGCAGTGCAGGGTG  1634

Qy    521  GlnThrValArgValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMet  540
Db   1635  CAGACGGGTCAGGGTACAGCTGCGTGACCCCCAGCGGGTCCAGTACCTGGTGGGTATG  1694

Qy    541  LeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaPro  560
Db   1695  CTTCGGCAGAGCCCCGCCCTGACCACCCTGGATCTCAGCGGCTGCCAACTGCCCGCCCCC  1754

Qy    561  MetValThrTyrLeuCysAlaValAlaLeuGlnHisGlnLysCysGlyLeuLeuThrLeuSer  580
Db   1755  ATGGTGACCTACCTGTGTGCKGTCCTGAGCACCAGGATGCGGCCTGCAGACCCTCAGT  1814

Qy    581  LeuAlaSerValGluLeuSerGluGlnSerGluLeuGlnGluLeuGlnAlaValLysArgAla  600
Db   1815  CTGGCCTCTGTGGAGCTGAGCGGAGCAGTGCATTCAGGCTGTGAAGAGAGCA  1874

Qy    601  LysProAspLeuValIleLeuThrHisProAlaLeuAspArgGlyHisProGlnProLysGlu  620
Db   1875  AAGCCGGATCTGGTCATCACACACCCCAGCGCTGGACGGGCCACCCCAAGCCTCCCCAAGGAA  1934

Qy    621  LeuLeuSerThrPhe  625
Db   1935  CTCATCCTCGGACCTTC  1949

RESULT 2
AX168123

LOCUS       AX168123              3365 bp   DNA   linear   PAT 03-JUL-2001
DEFINITION  Sequence 41 from Patent WO0142288.
ACCESSION   AX168123
VERSION     AX168123.1  GI:14597940;
KEYWORDS
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 3365)
  AUTHORS   Burford,N., Baughn.M.R., Au-Young,J., Yang,J., Lu,D.A. and Reddy,R.
  TITLE     G-protein coupled receptors
  JOURNAL   Patent: WO 0142288-A 41 14-JUN-2001;
            Incyte Genomics, Inc. (US)
FEATURES             Location/Qualifiers
     source          1..3365
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="Incyte ID No: 1499408CB1"

BASE COUNT      612 a    1106 c    1114 g     533 t
ORIGIN

Alignment Scores:
Pred. No.:                1.32e-235
Score:                    3218.00
Percent Similarity:       100.00%       Length:        3365
Best Local Similarity:    100.00%       Matches:       625
Query Match:                            Conservative:  0
DB:                       6             Mismatches:    0
                                        Indels:        0
                                        Gaps:          0

US-10-029-347-2 (1-625) x AX168123 (1-3365)

Qy      1  MetLeuAlaGlnProGlyLeuArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
Db    799  ATGCTGGCCCAGCCCGCACAGGCCCTTCATCCTGACGGGGCGGGACGAGCTGCCGGCG  858

Qy     21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheAlaAlaSerGlyAlaAlaArg  40
```

JOURNAL      Submitted (01-FEB-2002) Inflammation, Millennium Pharmaceuticals
             Inc., 45 Sidney Street, Cambridge, MA 02139, USA
FEATURES             Location/Qualifiers
     source          1..2679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p15"
     CDS             1..2679
                     /note="PYPAF5; PYRIN-containing nucleotide-binding
                     site/leucine-rich repeat family member"
                     /codon_start=1
                     /product="PYRIN-containing APAF1-like protein 5"
                     /protein_id="AAL87105.1"
                     /db_xref="GI:19387136"
                     /translation="MDQPEAPCSSTGPRLAVARELLAALELSQEQLKRFHKLRDV
                     GPDGRSIPWGRLERADAVDLAEQLAQPYGPEPALEVARKTLKRADARDVAAQLQERRL
                     QRLGLGSGTLLSVSEYKKYREHVLQLHARVKERNARSVKITKRFTKLLIAPESAAPE
                     EALGPAEPEPGRARSDTHTFNRLFRRDEEGRRPLTVVLQGPAGGKTMAAKKILYD
                     WAAGKLYQGQVDFAFFMPCGELLERPGTRSLADLILDQCPDRGAPVPQMLAQPQRLLF
                     ILDGADLPALGQPEAAPCTDPFEAASGARVLGGLLSKALLPTALLVTTRAAAPGRL
                     QGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERAERAYRFVKENETLFALCFVPFVCW
                     IVCTVLRQQLEIGRDLSRTSKTTTSVYLLFITSVLLFTTSVLSSAPVADGPRLQGDLRNLCRLAR
                     EGVLGRAQFAEKELEQLELRGSKVQFLFLSKKELPGVLETVTVQFDSQEFLAA
                     LSYLLEDGGVPRTAAGGVGTLLRGDAQPHSHLVLTRFLFGLLSAERMRDIERHFGCM
                     VSERVKQEALRWVQGQGQGCGVAPEVTEGAKGLEDTEEPEEEEGEEPNYPLELLYC
                     LYETQEDAFVRQALCRFPELALQRVRFCRMDVAVLSVCVRCCPAGQALRLISCRLVAA
                     QEKKKKSLGKRLQASLGGGSSQGTTKQLPASLLHPLFQAMTDPLCHLSSLTLSHCKL
                     PDAVCRDLSEALRAAPALTELGLLNNRLSEAGLRMLSEGLAWPQCRVCTVRVQLPDPQ
                     RGLQYLVGMLRQSPALTTLDLLSGCQLPAPMVTYLCAVLQHQGCGLQTLSLASVELSEQ
                     SLQELQAVKVRAKPDLVITHPALDGHPCPPKELISTF"

BASE COUNT     458 a   878 c   934 g   409 t
ORIGIN

Alignment Scores:
Pred. No.:                                    Length:          2679
AF479748                                      Matches:         625
Score:                    6.21e-235           Conservative:    0
Percent Similarity:       3207.5C             Mismatches:      0
Best Local Similarity:    99.84%              Indels:          1
Query Match:              99.84%              Gaps:            1
DB:                       99.67%
                          9

US-10-029-347-2 (1-625) x AF479748 (1-2679)

Qy     1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluProAla  20
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   799  ATGTTGGCCCAGCCGCAGCGGCTCCTCTTCATCCTGGACGGGGCGGGGCGGGGCGACGCTCGCCGGCG  858

Qy    21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg  40
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   859  CTGGGGGGGCCCGAGGCCGCCGGCCGGGGGCCCCTTGAGGCGGCCAGCGGCGGCGCGCGG  918

Qy    41  ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValThrThr  60
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   919  GTGCTAGGCGGGCTGCTGAGCAAGGCGCTGCTGCCCACGGCCCTCCTGGTGACCACG  978

Qy    61  ArgAlaAlaAlaProGlyArgLeuGlnArgLeuCysSerProGlnCysAlaGluVal  80
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   979  CGGGCGGCCGCCGGGCGCCTGCAGCGGCTGTGCAGCCCGCAGTGCGCGGAGGTG  1038

Qy    81  ArgGlyPheSerAspLysAspLysLysLysTyrPheTyrLysPhePheArgAspGluArg  100
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1039  CGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGAGG  1098

Qy   101  AlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe  120
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1099  AGGGGCCGAGCGGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC  1158

Qy   121  ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGluGlyArg  140
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1159  GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTCGAGCTCGG  1218

Db   1939  CAGCGAGTGCGCCTTCGCCCATGACGACGTGGCTGCTGCTCTGAGCTACTGCGTGAGGTGCTGC  1998

Qy   401  ProAlaGlyGlnArgLeuLeuSerCysArgLeuValAlaAlaGlnGluLysLys  420
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1999  CCTGCTGCGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGGAGAAGAAG  2058

Qy   421  LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGlyGlySerSerGlnThr  440
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2059  AAGAAGAGCCTGGGAAGCCGGCTCCAGGCCAGTTCTCAAGGCACC  2118

Qy   441  ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeu  460
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2119  ACAAAACAACTGCCAGCCTCCTTCTTCATCCACCTTTCAGGCAATGACTGACCACTG  2178

Qy   461  CysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAsp  480
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2179  TGCCATCTGAGCAGCCTCAGCTGTCCCACTGCAAACTCCCTGACGCGGCCTGCCGACAC  2238

Qy   481  LeuSerGluAlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuAsnAsnArg  500
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2239  CTTTCTGAGGCGCTGAGGGCCAGCCCCCGCGCTGACGGAGCTGGGCCTCCTCCACACAGG  2298

Qy   501  LeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgVal  520
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2299  CTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGGGCTAGCCTGGCCGCAGTGCAGGGTG  2358

Qy   521  GlnThrValArgValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMet  540
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2359  CAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGGTATG  2418

Qy   541  LeuArgGlnSerProAlaLeuThrThrLeuAspLeuLeuSerGlyCysGlnLeuProAla  560
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2419  CTTCGGCAGAGCCCGGCCCTGACCACCCTGGATCTGCTGAGCGGGTGCCAACTGCCGGCCCCC  2478

Qy   561  MetValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSer  580
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2479  ATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGGATGCGGGCTGCAGACCCTCAGT  2538

Qy   581  LeuAlaSerValGluLeuSerGluGlnSerLeuGlnGluLeuGlnAlaValLysArgAla  600
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2539  CTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAGGCTGTGAAGAGAGCA  2598

Qy   601  LysProAspLeuValIleThrHisProAlaLeuAspGlyHisProCysProProLysGlu  620
             |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2599  AAGCCGGATCTGGTCATCACACACCCAGCCTGGACGGCGCACCACCCCAACCTCCCAAGGAA  2658

Qy   621  LeuIleSerThrPhe  625
             |||||||||||||||
Db  2659  CTCATCTCGACCTTC  2673

RESULT 3
AF479748
LOCUS       AF479748                2679 bp    mRNA    linear    PRI 12-AUG-2002
DEFINITION  Homo sapiens PYRIN-containing APAF1-like protein 5 mRNA, complete
            cds.
ACCESSION   AF479748
VERSION     AF479748.1  GI:19387135
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 2679)
  AUTHORS   Wang,L., Manji,G.A., Grenier,J.M., Al-Garawi,A., Merriam,S.,
            Lora,J.M., Geddes,B.J., Briskin,M., DiStefano,P.S. and Bertin,J.
  TITLE     PYPAF7, a Novel PYRIN-containing Apaf1-like Protein That Regulates
            Activation of NF-kappa B and Caspase-1-dependent Cytokine
            Processing
  JOURNAL   J. Biol. Chem. 277 (33), 29874-29880 (2002)
   PUBMED   12019269
REFERENCE   2  (bases 1 to 2679)
  AUTHORS   Bertin,J.
  TITLE     Direct Submission

Qy  141  AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal 160
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1219 GACCTGTCCCGCACGAGCAGTCCAAGACCACCAGTGTACCTGCTTTCATCACCAGCGTT 1278

Qy  161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyArgLeuAspArgAsnLeuCys 180
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1279 CTGAGCTCCGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCCGACTGGACCGCAATCTGTGC 1338

Qy  181  ArgLeuAlaAArgGluGlyGluGlyGlyValLeuGlyValArgArgAlaGlnPheAlaGlyAlaGluLysGluLeuGlu 200
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1339 CGGCTGGCCGCCCGGGAGGGCGAGGGCGGTTGGTCGGACGGAGGAGCGCAGTTTGCCGAGAAGGAACTGGAG 1398

Qy  201  GlnLeuGlnLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLysGluLeuPro 220
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1399 CAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCTGCCG 1458

Qy  221  GlyValLeuGluThrGluValThrTyrGlnInPheIleAspGlnSerPheGlnGluPheLeu 240
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1459 GGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTTCCTC 1518

Qy  241  AlaAlaLeuSerTyrLeuLeuGluAspGlyValProArgThrAlaAlaGlyGlyVal 260
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1519 GCGGCACTGTCCTACCTGCTGGAGGACGGGGTGCCCAGGACCGCGGCTGGCGGCGTT 1578

Qy  261  GlyThrLeuLeuArgGlyAspAlaGlnInProHisSerHisLeuValLeuThrThrArgPhe 280
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1579 GGGACACTCCTGCGTGGGGACGCCCAGCCCCACAGCCACCTTGTGCTCACCACGCGCTTC 1638

Qy  281  LeuPheGlyLeuLeuSerArgAlaGluArgMetArgAspPheGluGlyArgHisPheGlyCysMet 300
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1639 CTCTTCGGACTGCTGAGCCGCGCGGAGCGCATGCGGGACTTCGAGGGCCGGCACTTCGGCTGCATG 1698

Qy  301  ValSerGluGluArgValLysGlnGluGluAlaLeuArgTrpValGlnGlyGlnGlyGlyGlyCys 320
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1699 GTTTCAGACGCTGTGAAGCAGGAGGAGGCGCTGCGGTGGGTGCAGGGACAGGGGTGC 1758

Qy  321  ProGlyValAlaProGluValThrGluGluGluGlyLeuGlyLeuGluAspAsnThrGluGluPro 340
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1759 CCCGGAGTGGCACCAGAGGTGACCGAGGAGGAGGGGCTGGGCCTGGAGGACAACACCGAGGAGCCA 1818

Qy  341  GlyGluGluGluGlyLyGlyGluProAsnGlyTyrProLeuGlnLeuLeuTyrCysLeuTyr 360
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1819 GGAGGAGGAGGAGGGGCCGAGCCCAATGGCTACCCACTACCCCTGCAGTTGCTGTACTGCCTGTAC 1878

Qy  361  GluThrGlnGluGluGluAspGlnAlaPheValArgGluHisLeuAsnGluValGluArgPheArgGlnLeuAspGluleu 380
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1879 GAGACGCAGGAGGAGGACCAGGCGTTTGTGCGCGAGCACCTGAACGAGGTGGAGCGGTTCCGACAGCTGGACGAGCTG 1938

Qy  381  GlnArgValArgPheCysSerArgMetAspAlaValLeuSerTyrCysValArgCysCys 400
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1939 CAGCGAGTGCGCTTCTGCAGCCGCATGGACGCGGTGCTGTCTTCTGAGCTACTGCCTGGAGCGTGCTGC 1996

Qy  401  ProAlaGlyValGlnArgAlaLeuLeuIleSerCysLysLysValAlaGlnGluValGlnLys 420
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1999 CCTGCTGGACAGGCACTGCGGCTGATTCAGCTGCAGATTGGTTGCTGCCAGGAGAAGAAG 2058

Qy  421  LysLysSerLeuGlyLysLysLysArgLeuGluGlnLeuAlaSerLeuGlyGly--SerSerGlnGlyGly 439
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2059 AAGAGAGCCTGGGGAAGCGGCTCGGGAAGAGGCGGCTGGAGCAGTTCTCAAGC 2118

Qy  440  ThrThrLysGlnLeuProAlaSerLeuHisProLeuPheGlnAlaMetThrAspPro 459
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2119 ACCACAAAACAACTGCCAGCCTCCTTCTCTTCATCCACCCTCTTCAGGCCAATGACTGACCCA 2178

Qy  460  LeuCysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArg 479
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2179 CTGTGCCACCTGAGCAGCCTTCACGCTGTCCCACTGCAAACTCCCTGACGCGTCGCCCA 2238

Qy  480  AspLeuSerGluAlaLeuAlaAlaProAlaLeuThrGluLeuGlyLeuLeuHisAsn 499
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2239 GACCTTTCTGAGGCAGAGCCTCGCGACGGAGCCTGGCCTTCTCCAAAC 2298

Qy  500  ArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGluGlyLeuAlaTrpProGlnCysArg 519

Qy  520  ValGlnThrValArgValGlnLeuLeuProAspProGlyArgGlyArgLeuGlnTyrLeuLeuValGlyLys 539
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2359 GTGCAGACGGTCAGGGTACAGCTGCCTGACCCCGGCGAGGGCCGAGGCTCCAGTACCTGGTGGT 2418

Qy  540  MetLeuArgGlnInSerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAla 559
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2419 ATGCTTCGGCAGCCCTGCCCCTGACCACCCTGGATCTCAGCGGGCTGCCAACTGCCCGCC 2478

Qy  560  ProMetValThrTyrLeuCysValAlaValAlaValLeuGlnInHisGlnInGlyLeuGlyLeuCysGlyGlyLeuThrLeu 579
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2479 CCCATGGTGACCTACCTGTGCGTGGCTGTGGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGGGCCTC 2538

Qy  580  SerLeuAlaSerValGluGluSerGluGlnInSerLeuLeuGlnInLeuGlnInAlaValLysArg 599
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2539 AGTCTGGCCTCTGTGGAGGAGAGTGAGCAGAGCTTCAGGCTGCTGCAGGCTGCTGTGAAGAGA 2598

Qy  600  AlaLysProAspLeuValIleThrHisProAlaLeuAspGlyHisProGlnInProProLys 619
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2599 GCAAAGCCGGATCTGGTCATCACACACCCAGGCTGCACGCGGGCCACCCCACAACCTCCCAAG 2658

Qy  620  GluLeuIleSerThrPhe 625
         |||||||||||||||||
Db  2659 GAACTCATCTCGACCTTC 2676

RESULT 4
AY:5446-

LOCUS        AY154461              2679 bp    mRNA    linear   PRI 20-FEB-2003
DEFINITION   Homo sapiens NALP6 (NALP6) mRNA, complete cds.
ACCESSION    AY154461
VERSION      AY154461.1  GI:28436365
KEYWORDS     .
SOURCE       Homo sapiens (human)
  ORGANISM   Homo sapiens
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE    1  (bases 1 to 2679)
  AUTHORS    Tschopp,J., Martinon,F. and Burns,K.
  TITLE      NALPs: a novel protein family involved in inflamation
  JOURNAL    Nat. Rev. Mol. Cell Biol. 4 (2), 95-104 (2003)
  MEDLINE    22451042
   PUBMED    12563287
REFERENCE    2  (bases 1 to 2679)
  AUTHORS    Martinon,F., Hofmann,K. and Tschopp,J.
  TITLE      Direct Submission
  JOURNAL    Submitted (25-SEP-2002) Institute of Biochemistry, University of
             Lausanne, ch. les Boveresses 155, Epalinges, VD 1066, Switzerland
FEATURES             Location/Qualifiers
     source          1..2679
                     /organism="Homo sapiens"
                     /mol_type="mRNA"
                     /db_xref="taxon:9606"
     gene            1..2679
                     /gene="NALP6"
     CDS             1..2679
                     /gene="NALP6"
                     /note="member of the NALP protein family involved in
                     inflammation; contains NACHT, LRR and PYD domains"
                     /codon_start=1
                     /product="NALP6"
                     /protein_id="AAO18157.1"
                     /db_xref="GI:28436366"
                     /translation="MDQPRAPCSSTGPRLAVARELLAALEELSOEQLKRFRHKLRDV
                     GPDGRSIPWGRLBRADAVDLAEQLAPQFYGPEPALEVARKTLKRADARDVAAQLOERRL
                     QRLGLGSGTLLSVSEYKKKYREHVLQLHARVKERNARSVKITKRFTKLLIAPESAAPE
                     EALGPABEEPEPGRARRSDTHTFNRLPRRDEBGRRPLTVVLQGRGAPVPQMLAQPQRLLF
                     WAAGKLYOGQVDPAFFMPCGELLBRPGTRSLADLILDQCPDRGAPVPQMLAQPQRLLF
                     ILDGADELPALGGPEEAAPCTDPFEAASGARVLGGLLSKALLPTALLVTTRAAAPGRL
                     QGRLCSPQCAEVRGFSDKDKKKYFYKFFRDERRAERAYRFVKENETLFALCFVPFVCW
                     IVCTVLRQLELGRDLSRTSKTTSVYLLFITSVLSSAPVADGPRLQGDLRNLCRLAR
                     EGVLGRRAQFAEKELEQLERGSKVQTLFLSKKELPGVLSTEVTYQFIDQSFQEFLAA"

<!-- Right column top segment (Qy/Db alignment continuation) -->
Qy  520  ValGlnThrValArgValGlnLeuLeuProAspProGlyArgGlyArgLeuGlnTyrLeuLeuValGlyLys 539
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2299 AGGCTCAGTGAGGCGGGACTGCGTATGCTGAGTGAGGAGGGCCTAGCCTGGCCGCAGTGCAGG 2358

LSYLLEDGVPRTAAGGVGTLLRGDAQPHSHLVLTTRFGLLSAERMRDIERHFGCM
VSERVKQEALRWVQGQGQCPGVAPEVTEGAKGLEDTEPEEEEGEEPNYPELELYC
LYETQEDAFVRQALCRFPELALQVRFCRMDVAVLSYCVRCCPAGQALRLISCRLVAA
QEKKKSLGKRLQASLGGGSSGSQGTTKQLPASLLHLPFQAMTDPLCHSSSLTLSHCKL
PDAVCRDLSEALRAAPALTELGLLHNRLSEAGLRMLSEGLAWPQCRVQTVRVCLPDPQ
RGLDYILVGMLRQSPALTTLDLSGCQLPAPMVTYLCAV.QHQGCGLQT.SLASVELSEC
SLQEELQAVKRAKPDLVITHPALDGHPQPPKELISTF"

BASE COUNT    458 a     878 c     934 g     409 t
ORIG.N

Alignment Scores:
Pred. No.:          6.21e-235
Score:              3207.50
Percent Similarity: 99.84%
Best Local Similarity: 99.84%
Query Match:        99.67%
DB:                 9

US-10-C29-347-2 (1-625) x AY154461 (1-2679)

```
Db      2659 GAACTCATCTGACCTTC 2676

RESULT 5
AX663298
LOCUS       AX663298               2763 bp    DNA        linear   PAT 22-MAR-2003
DEFINITION  Sequence 5 from Patent WO2061086.
ACCESSION   AX663298
VERSION     AX663298.1  GI:29163662
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Feder,J., Ramanathan,C. and Mintier,G.
  TITLE     Human leucine-rich repeat containing protein, expressed
            predominantly in small intestine, HLRRSI1
  JOURNAL   Patent: WO 02061086-A 5 08-AUG-2002;
            Bristol-Myers Squibb Company (US)
FEATURES             Location/Qualifiers
     source          1..2763
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT      544 a    890 c    970 g    459 t
ORIGIN

Alignment Scores:
Pred. No.:                          Length:         2763
Score:          4.67e-232           Matches:        624
Percent Similarity:  3170.00        Conservative:   0
Best Local Similarity:  96.00%      Mismatches:     1
Query Match:    96.00%              Indels:         26
DB:             98.51%              Gaps:           1
                6

US-10-029-347-2 (1-625) x AX663298 (1-2763)
```

```
Qy     1  MetLeuAlaGlnProGlnArgLeuPheIleLeuLeuAspGlyAlaAlaAspGluLeuProAla 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75  ATGCTGGCCCAGCCGCAGCGGCTGTTCATCCTGCTGGATGGCGCGGCCGCCGAGCTGCCGGCG 134

Qy    21  LeuGlyGlyProGluAlaAlaAlaProCysPheAspAspProPheGluAlaAlaSerGlyAlaAlaArg 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135  CTGGGGGGCCCCGAGGCCGCGGCCCCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCGGCGCGG 194

Qy    41  ValLeuGlyGlyLeuLeuSerLysLysAlaLeuAlaLeuProThrAlaLeuLeuLeuValThrThr 60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195  GTGCTAGGCGGCGGCTCGCTGAGTAAGGCGCTGGCCCTGCCCACGGCCCTCCTGGTGGTGACCACG 254

Qy    61  ArgAlaAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnGlyArg 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255  CGCGCCGCGGCCGCCCCGGGGAGGCTGCAGGGCCGCCTGTGTTCCCCGCAGGGGCGCGAGGTG 314

Qy    81  ArgGlyPheSerAspLysAspLysLysLysTyrPheTyrLysSerPhePheArgArgGluArg 100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315  CGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTCTTTCCGGGATGAGAGG 374

Qy   101  ArgAlaGluArgAlaTyrArgPheValLysThrValGlnAsnGluThrLeuPheAlaLeuCysPro 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   375  AGGGCCGAGCGCGCCTACCGCTTCGTGAAGACAGCGCAGAACGAGACGCTGTTCGCGCTGTGCTTC 434

Qy   121  ValProPheValCysThrValIleValTrpIleValGlnLeuGluLeuGlyArg 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   435  GTGCCCTTCGTGTGCACCGTGCACCGTGCACCGTGGCACCGTGCGCCAGCTGGAGCTCGGTCGG 494

Qy   141  AspLeuSerArgThrSerLysThrThrThrValTyrLeuGluPheIleThrSerVal 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   495  GACCTGTCGCGCACGTCCAAGACCACCACGGTACAGTCAGTGTACCTGGGAGTTCATCACCAGCGTT 554

Qy   161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCys 180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   555  CTGAGCTCGGCGCCCGTAGCCGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGC 614
```

```
Qy   181  ArgLeuAlaArgGluArgGlyValLeuGlyArgArgAlaGlnPheAlaGlyGluLeuGlu 200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   615  CGCCTGGCCCGCGAGCGGGGCGTCCTCGGACGGCGGGCGCAGTTCGCCGGAGAAGAACTGGAG 674

Qy   201  GlnLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeuSerLysLysGluLeuPro 220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   675  CAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTCTCAGCAAAAAGGAGCTGCCG 734

Qy   221  GlyValLeuGluThrGluValThrTyrGlnPheIleAspGlnSerPheGlnGluPheLeu 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   735  GGGGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGGAGTTCCTC 794

Qy   241  AlaAlaLeuSerTyrLeuLeuGluLeuGluAspGlyValGlyValProAlaArgThrAlaAlaGlyValVal 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   795  GCGGCACTGTCCTACCTGCTGGAGCTGGAGGACGGGGTGGGCGTGCCCGGCTCGGCGCCGGGGCGGTT 854

Qy   261  GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrThrArgPhe 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   855  GGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACCTGGTGCTCACCACGCGCTTC 914

Qy   281  LeuPheGlyLeuLeuSerLeuAlaGluArgMetArgAspPheGluArgHisPheGlyCysMet 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   915  CTCTTCGGACTGCTGAGCCTGGCCGAGCGCATGCGCGACTTCGAGCGCCAC-TTCGGCTGCATG 974

Qy   301  ValSerGluValLysGlnValGlnLeuAlaLeuArgTrpValGlnGlyLy.nGlyGlnGlyCys 320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   975  GTTTCAGAGCGTGTGAAGCAGGTGCAGCTGGCCCTGCGCTGGGTGCAGGGACAGGGCGGCTGC 1034

Qy   321  ProGlyValAlaProGluValThrGlyLysLeuGluAspThrGluPro 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1035  CCCGGAGTGGCACCAGAGGTGACCGAGGGGCCCAAAGGGCTGGACACCGAAGAGCCA 1094

Qy   341  GluGluGluGluGluGluGluGluGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyr 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1095  GAGGAGGAGGAGGAGGAGGAGGAGCCCAACTACCCACCTGGAGTTGCTGTACTGCCTGTAC 1154

Qy   361  GluThrGlnGluAspAlaPheValAlaArgGlnAlaLeuAlaLeuAlaLeu 380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1155  GAGACGCAGGAGGACGCGTTTGTGGCGCAAGCCCTGGCCCTGGCGCTG 1214

Qy   381  GlyArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysValCys 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1215  CAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGCGCTGCGTGTGC 1274

Qy   401  ProAlaGlyGlnAlaLeuArgLeuIleSerCysArgGluValAlaAlaGlnGluLysLys 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1275  CCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGGGAGGTGGCTGCGCAGGAGAAGAAG 1334

Qy   421  LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGlySerGlnGlyThr 440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1335  AAGAAGAGCCTGGGAAAGCGGCTCCAGGCCAGCCTGGGTGGGGGCAGTTCTCAAGGACC 1394

Qy   441  ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuArg 460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1395  ACAAAACAACTGCCAGCCTCCCTTCTTCATCACCACTGACCCACTG 1454

Qy   461  CysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAsp 480
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1455  TGCCATTCTGAGCAGCCTCACGCTGTCCCACTGACCACTCCCTGACGGGCGTCTGCCAGAC 1514

Qy   481  LeuSerGluAlaAlaLeuArgAlaAlaProAlaLeuThrGlyLeuGlyLeuHisAsnArg 500
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1515  CTTTCTGAGGCCCGAGGCAGCCCACTGACGGACGTGGCCTCCTCCCACACAGG 1574

Qy   501  LeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnArgGlyVal 520
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1575  CTCAGTGAGGCGGGACCTGCGTATGCTGAGTGAGGGCCTAGCCTGGCCCCAGTGCAGGGTG 1634

Qy   521  GlnThrVal------------------------- 523
          |||||||
Db  1635  CAGACGGT-CAGGTGAGGCCTGCCTGGGAGGGACCCGGGATGCCCCGCCACCCCAGC 1693

Qy   524  ------ArgValGlnLeuProAspProGlnArgGlyLeuGlyArgLeuGluGln- 535
```

RESULT 6
AX088215

```
LOCUS       AX088215                5553 bp    DNA        linear    PAT 17-MAR-2001
DEFINITION  Sequence 10 from Patent WO0114564.
ACCESSION   AX088215
VERSION     AX088215.1  GI:13397126
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Feyman,J.A., Green,C.E., Hsu,A., Browning,J.A. and Carulli,J.
  TITLE     Polynucleotides expressed in activated t-lymphocytes and proteins
            encoded thereby
  JOURNAL   Patent: WO 0114564-A 10 01-MAR-2001;
            Curagen Corporation (US) ; BIOGEN, INC. (US)
FEATURES             Location/Qualifiers
     source          1..5553
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
BASE COUNT        912 a    1856 c    1879 g     906 t
ORIGIN
```

US-10-029-347-2 (1-625) x AX088215 (1-5553)

Alignment Scores:
Pred. No.:
Score:                    2882.00
Percent Similarity:       82.82%
Best Local Similarity:    81.83%
Query Match:              89.56%
DB:

Matches:       5553
               581
Conservative:  7
Mismatches:    32
Indels:        90
Gaps:          5

RESULT 7
AX459875

LOCUS       AX459875          5556 bp    DNA     linear   PAT 08-JUL-2002
DEFINITION  Sequence 18 from Patent WO0240668.
ACCESSION   AX459875
VERSION     AX459875.1  GI:21725646
KEYWORDS    .
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
  AUTHORS   Tschopp,J. and Martinon,F.
  TITLE     Proteins and dna sequences underlying these proteins used for
            treating inflammation
  JOURNAL   Patent: WO 0240668-A 18 23-MAY-2002;
            Apotech Research and Development Ltd. (CH)

FEATURES             Location/Qualifiers
     source          1..5556
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /note="NALP6/Pyp.dna"

BASE COUNT      914 a   1856 c   1879 g    907 t
ORIGIN

Alignment Scores:
Pred. No.:
Score:                                  1.03e-209
                                        2882.00              Length:        5556
Percent Similarity:     82.82%          Matches:       581
Best Local Similarity:  81.83%          Conservative:    7
Query Match:            89.56%          Mismatches:     32
Db:                            6                              Indels:         90
                                                             Gaps:            5

US-10-029-347-2 (1-625) x AX459875 (1-5556)

```
Qy    583  ----------------------------------------SerValGluLeuSerGlu----------------  588
                                                  |||           |||
Db   2719  TCTGACCCGCACCCGAGTTCTTCTCCGGACGTGTGCAGAACCCGGCGCCGGCGCCTGGGG  2778

Qy    589  --------------------GlnSerLeuGlnGluLeuGlnLeuAlaValLysArgAlaAlaLysPro  602
                               :::|||
Db   2779  CTGGAGTCTCGGTCGCCTCGGAGCGCCCCGAGCCCTCGGGCCACAGCGAGCGGAGGACC  2838

Qy    603  AspLeuValIleThr----------------------------------HisProAla  610
                :::
Db   2839  CAGGTGGAGGCGGCGGCGGCGGCGGCGGGAGGAGGAGGCGCGGGAGGGAACCCCCGGCC  2898

Qy    611  LeuAspGlyHisProGlnProProLysGlu  620
                |||||||||||||||||||
Db   2899  CGCGGGCGCCCCACCCACAGCGCGCCCGGAC  2928
```

RESULT 8
AX088208
LOCUS       AX088208                    6461 bp    DNA        linear        PAT 17-MAR-2001
DEFINITION  Sequence 3 from Patent WO0114564.
ACCESSION   AX088208
VERSION     AX088208.1  GI:13397119
KEYWORDS    .
SOURCE      Homo sapiens (human)
ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1
AUTHORS   Peyman,J.A., Green,C.D., Hsu,A., Browning,J.A. and Carulli,J.
TITLE     Polynucleotides expressed in activated t-lymphocytes and proteins
            encoded thereby
JOURNAL   Patent: WO 0114564-A 3 01-MAR-2001;
            Curagen Corporation (US) ; BIOGEN, INC.  (US)
FEATURES             Location/Qualifiers
source          1..6461
                  /organism="Homo sapiens"
                  /mol_type="genomic DNA"
                  /db_xref="taxon:9606"
CDS             1..5556
                  /note="unnamed protein product"
                  /codon_start=1
                  /protein_id="CAC34689.1"
                  /db_xref="GI:13397120"
                  /translation="MDQPEAPCSSTGPRLAVARELLLAALELSQPCLKRFRHKLRDV
                  GPDGRSTPWGRLERADAVDLAEQLAQFYGPEPALEVARKTLKRADRDVAAQLQERR..."

```
RGLLDRSFANMAEPFKVWTENADGSGAVNFLTGMGGFLQAVVFGCTGFRVTRAGVTFD
PVCLSGISRVSVSJIFYOGNKLNFSFEDSVTVEVTARAGPWAPHLEAELMPSQSRLS
LLPGHKVSFPRSARRIOMSPPPKLPGSSSSSEFPGRTFSDVRDPLQSPLWVTLGSSSPTE
SLTVDPASE"
```

```
BASE COUNT   1079 a   2146 c   2150 g   1086 t
ORIGIN
```

Alignment Scores:
```
Pred. No.:           1.25e-209    Length:       6461
Score:               2882.00      Matches:      581
Percent Similarity:  82.82%       Conservative: 7
Best Local Similarity: 81.83%     Mismatches:   32
Query Match:         89.56%       Indels:       90
DE:                               Gaps:         5
```

US-10-029-347-2 (1-625) x AX088208 (1-6461)

```
Qy    1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla    20
          || |  | | |  ||| ||  ||  |  |  || |  ||||||||||| ||| ||
Db  799 ATGCTGGCCCAGCCGCTGCTCTTCATCCTGGACGGCGCCGGAGAGCTGCCGGCG       858

Qy   21 LeuGlyGlyProGluAlaAlaProPheGluAlaAlaSerGlyAlaArg              40
          || | |||||| |  | ||  ||  ||  || ||  | |||| |
Db  859 CTGGGGGGCCCCGAGGCCGCGCCCTTCGACAGACCCCTTCGCAGGGCCGAGGCGGCGG    918

Qy   41 ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr    60
          || |  | | | |  ||| |  |  |  ||| |  ||| ||  |  |  | |  ||
Db  919 GTGCTAGGCGGGCTGCTGAGCAAGGCGCTGCTGCCCACGGCCCTCCTGCTGGTGACCACG    978

Qy   61 ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuArgCysSerProGlnCysAlaGluVal    80
          |||||||| | |  |  ||| |  ||| || |  |  |  || |||
Db  979 CGCGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCCTGCGGTGTTCCCCGCAGTGCGCCGAGGTG   1038

Qy   81 ArgGlyPheSerAspLysAspLysLysLysTyrTyrPhePheLysPheAlaLysArg       100
          ||| | || |  ||  |  |  |  |  |  || |  ||  |  ||  | ||
Db 1039 CGGGGCTTCTCGGACAAGGACAAGAAGAAGTATTATTTCTTCAAGTTCGCCAAGAGG       1098

Qy  101 ArgAlaGlnArgProPheValLysGluGluAsnGluThrAlaLeuCysPhe          120
          || || |  || | || |   | ||  || |  | |  | ||  || |
Db 1099 AGGGCCCAGCGGCCGTTCGTGAAGGAGGAGAACGAGACGGCTCTGTGCTTC          1158

Qy  121 ValProPheValCysTrpLeuValCysThrValLeuArgGlnLeuLeuGluLeuGlyArg    140
          | |  | | ||| |  | | |  |  |||| || |  ||||||  ||  | |||
Db 1159 GTGCCCTTCGTGTGCTGGCTGGTGTGCACCGTGCTCCGGCAGCTGCTGGAGCTGGGTCGG   1218

Qy  141 AspLeuSerArgThrSerLysThrThrValTyrLeuLeuThrLeuLeuSerVal         160
          |  ||| || | || |  |  |  | |  || | |  |  || ||
Db 1219 GACCTGTCGCGCACCTCCAAGACCACAGTGTACCTGCTTTTCATCTTCATCACCAGCGTT    1278

Qy  161 LeuSerSerAlaProValAlaAspGlyProArgLeuGlnCysTyrAspLeuArgAsnLeuCys    180
          ||  || |||| | | |  | | |  | ||| |   ||  | ||| | ||
Db 1279 CTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGCTGCAGGGCTACGACCTGCGGAATCTGTGC    1338

Qy  181 ArgLeuAlaArgGlyGlyValLeuGlyLeuArgArgAlaGlnPheAlaLeuLysGlyLeuGlu    200
          ||| | ||| | ||  | | | || | |  ||||||  ||| |||  |||  |
Db 1339 CGCCTGGCCCGCGGGGGCGTCCTGGACCTGCGGCGCGCACAGTTTGCCCTGAAGGGACTGGAG   1398

Qy  201 GlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLeuLysLysLysGluPro    220
          || || |  ||| |  |  |  | | |  || ||  | ||  | |  | || || | ||
Db 1399 CAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCTGCCG    1458

Qy  221 GlyValLeuGluThrGlnValThrTyrArgGlnSerProGlnSerPheGlnLeuPheLeu    240
          | | || ||  | |  | | |  | | ||| |  ||| |   ||  | |||
Db 1459 GGCGTGCTGGAGACGCAGGTGACCTACCGGCAGTCTCCGCAGAGCTTCCAGCTGTTCCTC    1518

Qy  241 AlaAlaLeuSerTyrLeuGluLeuGluAspGlyValProArgThrAlaAlaGlyValGlyVal    260
          |||  | || | |  || |  ||  ||||| |  ||| |  |  |  | |  | | |
Db 1519 GCCGCACTGTCTCTACCTGTCGGAGGACGGCGTGCCCAGGACCGCCGCTGGCGGGCGTT    1578

Qy  261 GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisValLeuThrThrArgPhe       280
          | | || |  || | || | |  || |  || | |  ||  | | | || ||
Db 1579 GGGACACTCCTGCGTGGGGACGCCCAGCCCCACAGCCACGTGCTGCTCACCACCAGGTTC    1638
```

```
Qy  281 LeuPheGlyLeuLeuSerAlaGlnMetGlyGlyCysMet                        300
          |  | || |  |  | |||  | |||  | | ||||  |
Db 1639 CTCTTCGGACTGCTGAGCGCGGAGCACGGCGATGCGCGACATCGAGCCCACTTCGGCTGCACG   1698

Qy  301 ValSerGluArgValLysGlnGluValGlyAlaLeuArgTrpValGlnGlyValGlnGlyCys   320
          || || || | | | |  |  | ||  | || ||| ||  |  |  || ||  |  |
Db 1699 GTTTCAGACGGTGGAGCAGGAGGTGGGTGCAGGGACAGCAGGGACAGGGTGC      1758

Qy  321 ProGlyValAlaProGluValThrGlyGlyValThrGlyLeuGluAspThrGluGlyLeuPro   340
          ||  | || || ||  | |  | | | | |  |  | |  || || |  || | || ||
Db 1759 CCCGGAGTGGCACCCAGAGGTGACCGGAGGGGGCAAAGGGCTCGAGGACACCGAAGGGCCA    1818

Qy  341 GlyGluGluGluGlyGluGluGluGluGluGluGluGluGluGluGluProAsnTyrCysLeuTyr   360
          | || || ||  | || ||  | || || || || || || |   || |  || |
Db 1819 GGAGAAGAAGAGGGAGAGGAGGAGGAAGAGGAGGAGGAGCCCAACTACTGCCTGTAC    1878

Qy  361 GluThrGluAspAlaPheGlyValArgGlnAlaLeuCysSerArgPheProGluLeuAlaLeu   380
          | |  || || | || | | | ||  |||  |  ||| || ||  || || | |
Db 1879 GAGACGGAGGACGCGTTTGGCGCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCCCTG    1938

Qy  381 GlnArgValArgPheCysSerArgMetAspValArgValAlaValLeuSerTyrArgCysValArgCysCys   400
          || || | || | || || || | ||| |  |  | | |  | || || |  || |  | |||
Db 1939 CAGCGAGTGCGCTTCTGCCGCATGGACGTGCGCGTGGCTGTTCTGAGCTACCGCTGCGTGCGCTGCTGC    1998

Qy  401 ProAlaGlyGlnAlaLeuAlaArgLeuIleSerCysArgLeuValAlaLeuAlaGlnLeuLysLys   420
          ||  | |  || |  ||| | ||| || | || ||| | || |  ||| |  | || | ||
Db 1999 CCTGCTGGACAGGCCCTGGCTCGAGCTGATCAGCTGCAGGCTGGTTGCTGCCGCAGAAGAAG    2058

Qy  421 LysSerLeuGlyLysArgLeuGlyLeuAlaSerLeuGlyLyGlySerArgSerGlnGlyThr   440
          | ||| | | || |  || || | |||| |  || | | | |  ||  |||  || ||
Db 2059 AAGAGAGCCTGGGGAAGCGGCTTCCAGGCCAGCCTGGGTGGCGGCCAGTGGCTGGCCCAGTGGGGACC    2118

Qy  440 --------------------------------------------------------        440
Db 2119 CAACTGGCTCCAGAAGTACCCTTCGACAACCTCGTGTGACATCGTGCCCCACACCTCCA       2178

Qy  440 --------------------------------------------------------        440
Db 2179 CCAGACCCCTGCCTCCTCCAGGCGCAAGGCTTTGCCCAGAGTTCCTTTGAATATAGTCCA    2238

Qy  441 ThrLysGlnLeuProAlaSerLeuLeu---                                  449
          || | |  ||  |  | |  |  |
Db 2239 ATTCAGCCCTGCCCAGGGCTTGGCATCTGTGGCATCGTGTGAGAGATGAATGTCACGGTGTTGGCA    2295

Qy  450 ------HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSer             464
                |  ||| |  || |  ||| | | || ||  || ||  |
Db 2299 GGGGCTGGCGTTGGGACCCAAAGACCCCATGAAGATGACTGACCCACTGTGCCATCTGAGC    2358

Qy  465 SerLeuThrLeuSerHisCysLysLysLeuProAspAlaValCysArgAspLeuSerGlnGluAla    484
          || | || |  || |  | | | ||  || | | | | ||| |  || | || |
Db 2359 AGCCTCACGCTGTCCCACTGCAAACTCCCTGACGCCGTGTGCCGGGACCTTTCTGAGGCC    2418

Qy  485 LeuArgAlaAlaProAlaLeuLeuThrGlyLeuGluHisAsnArgLeuSerGlnValAla    504
          ||| |||  | ||  || |  | |  | || |  | | ||| |  || ||  | |
Db 2419 CTGAGGGCAGCCCCCGCACTGCTGACGGGCCTGGAGCACAACAGGCTCAGTGAGGCA    2478

Qy  505 GlyLeuArgMetLeuSerGluGlyLysValGlyLnThrValArgGlyLeuArgValGlnThrValArg   524
          | || ||| |  | |  || | | |  || ||  | || |  | || | | ||  |||
Db 2479 GGACTGCGTATGCTGAGTGAGGGCGCCAGTGCCCAGGGCGCAGACGGGTCAGG    2538

Qy  525 ValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSer    544
          | || |  ||  || |  |  | | | || |  | | | | | |  || | ||
Db 2539 GTACAGCTGCCTGACCCCCCAGCGGGGGCTCCAGTACCTGGTGGGTATGTTGCGGCAGAGC    2598

Qy  545 ProAlaLeuThrThrLeuAspLeuSerGlyLeuCysGlnLeuProAlaProMetValThrTyr    564
          ||  | |  | | |  || || ||  | |  ||| |  || | | | ||  | ||
Db 2599 CCTGCCCTGACCACCCTGGATCTCAGCGGGCTGTGCCAACTGCCCGCCCCCATGGTGACCTAC    2658

Qy  565 LeuCysAlaValLeuGlnHisGlnGlyLeuGlnThrLeuSerLeuLeuAla------       582
          |  || |  | | || || | || || | || ||  |  |  |
Db 2659 CTGTGTGCAGTCCTGCAGCACCAGGGCCTCCAGACCCTCAGCCTCCTGGCTCCT       2718

Qy  583 ------SerValGluLeuSerGlu                                        588
```

```
Db  2719 TCTGACCCGACCCGAGTTCTTCTCCGAGGGTCGAGAACCCGGCGCCGGCTCGGG 2772
                |||                                        |||
Qy   589 --------------GlnSerLeuGlnGluLeuGlyLeuAlaValLysArgAlaLysPro 602
                :||||||                                 |||
Db  2779 CTGGAGTCTCGCTGGCCTCGGAGCGCGCCCCAGCCCTCGGCGACAGGGAGGAGGACC 2838
                                    |||:|||
Qy   603 AspLeuValIleThr---------------------------HisProAla 620
                :|||                                        ||||||
Db  2839 CAGTGAGGAGGCGGCCGGGGGCGGCGGCGGCGGAGGAGGCGCGGGAGGGAACCCCCGCC 2896
                |||||||                    |||:|||:||
Qy   611 LeuAspGlyHisProGlnProProLysGlu 620
                |||||||:||||||:||:::::
Db  2899 CGCGGGCCCCACCCACAGCCGCCCCGGAC 2928

RESULT 9
AF015416
LOCUS       AF015416               95038 bp    DNA     linear   PRI 20-AUG-1997
DEFINITION  Homo sapiens chromosome 11 from 11p15.5 region, complete sequence.
ACCESSION   AF015416
VERSION     AF015416.1  GI:2335202
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
ORGANISM    Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 95038)
AUTHORS     Evans,G.A., Bradbury,F., Brignac,S., Burmeister,R., Davie,J.,
            Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
            Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,C.,
            Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
            Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
            Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
            Wilson,R. and Burbee,D.
TITLE       95 Kb from Four Overlapping Human Chromosome 11p15.5 Cosmids
JOURNAL     Unpublished
REFERENCE   2  (bases 1 to 95038)
AUTHORS     Evans,G.A., Bradbury,F., Brignac,S., Burmeister,R., Davie,J.,
            Davies,C.J., Davis,C., English,C., Fondon,T., Franklin,T.L.,
            Garner,H.R., Gordon,M., Gotway,G., Grant,O., Hahner,L., Harris,C.,
            Hinson,S., Megarity,C., Narayanaswamy,U., Newton,J., O'Brien,K.,
            Oliver,T., Patel,P., Probst,S., Rayner,S., Schageman,J.,
            Schilling,P., Schultz,R., Syed,M., Valenzuela,D., Ward,T.,
            Wilson,R. and Burbee,D.
TITLE       Direct Submission
JOURNAL     Submitted (18-JUL-1997) Genome Science and Technology Center,
            University of Texas Southwestern Medical Center atDallas, 5323
            Harry Hines Blvd, Dallas, TX 75235-8591, USA
FEATURES             Location/Qualifiers
     source          1..95038
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11p15.5"
     repeat_region   711..975
                     /rpt_family="Alu"
     repeat_region   2728..3022
                     /rpt_family="Alu"
     repeat_region   3363..3729
                     /rpt_family="Alu"
     repeat_region   5634..5909
                     /rpt_family="THE"
     repeat_region   complement(6220..6959)
                     /rpt_family="Alu"
     repeat_region   10560..10839
                     /rpt_family="Alu"
     repeat_region   complement(11832..12455)
                     /rpt_family="Alu"
     repeat_region   12921..13405
                     /rpt_family="L1"
     repeat_region   13778..14314
                     /rpt_family="Alu"
     repeat_region   14617..14779
                     /rpt_family="Alu"
     repeat_region   15396..15695
                     /rpt_family="Alu"
     repeat_region   15722..15994
                     /rpt_family="Alu"
     repeat_region   complement(16696..16997)
                     /rpt_family="Alu"
     repeat_region   complement(18996..19286)
                     /rpt_family="Alu"
     repeat_region   complement(19458..19747)
                     /rpt_family="Alu"
     repeat_region   complement(19777..20045)
                     /rpt_family="Alu"
     repeat_region   complement(22885..23002)
                     /rpt_family="MIR"
     repeat_region   23109..23261
                     /rpt_family="Alu"
     repeat_region   23292..23481
                     /rpt_family="MER46"
     repeat_region   complement(23877..23944)
                     /rpt_family="MIR"
     repeat_region   complement(24695..24986)
                     /rpt_family="Alu"
     repeat_region   complement(25429..26012)
                     /rpt_family="Alu"
     repeat_region   complement(27380..27676)
                     /rpt_family="Alu"
     repeat_region   complement(29422..29486)
                     /rpt_family="MER42"
     repeat_region   complement(29514..29802)
                     /rpt_family="Alu"
     repeat_region   30797..31075
                     /rpt_family="Alu"
     repeat_region   31688..31792
                     /rpt_family="Alu"
     repeat_region   32125..32412
                     /rpt_family="Alu"
     repeat_region   complement(38871..39186)
                     /rpt_family="Alu"
     repeat_region   42952..43094
                     /rpt_family="Alu"
     repeat_region   43106..43388
                     /rpt_family="MER5"
     repeat_region   43754..44216
                     /rpt_family="Alu"
     repeat_region   complement(44815..45136)
                     /rpt_family="MLT2B2"
     repeat_region   45145..45484
                     /rpt_family="MER42"
     repeat_region   45740..45995
                     /rpt_family="Alu"
     repeat_region   46375..46646
                     /rpt_family="Alu"
     repeat_region   47145..47738
                     /rpt_family="Alu"
     repeat_region   48353..48626
                     /rpt_family="Alu"
     repeat_region   49567..49842
                     /rpt_family="Alu"
     repeat_region   49940..50211
                     /rpt_family="Alu"
     repeat_region   complement(50615..50907)
                     /rpt_family="Alu"
     repeat_region   complement(53001..53290)
                     /rpt_family="Alu"
     repeat_region   complement(54161..54512)
                     /rpt_family="Alu"
     repeat_region   54533..54823
                     /rpt_family="L1"
     repeat_region   57899..58113
                     /rpt_family="Alu"
     repeat_region   complement(58795..59044)
                     /rpt_family="MER42"
```

repeat_region   /rpt_family="Alu"   complement(59414..60023)
repeat_region   /rpt_family="L1"   complement(60177..60311)
repeat_region   /rpt_family="Alu"   complement(60770..61108)
repeat_region   /rpt_family="L1"   complement(61207..61557)
repeat_region   /rpt_family="THE1"   61670..61942
repeat_region   /rpt_family="Alu"   61944..62436
repeat_region   /rpt_family="MER41"   complement(67807..68105)
repeat_region   /rpt_family="Alu"   70961..71049
repeat_region   /rpt_family="Alu"   71224..71523
repeat_region   /rpt_family="MER4"   complement(85033..85397)
repeat_region   /rpt_family="Alu"   85414..85620)
repeat_region   /rpt_family="MER4"   complement(85701..85804)
repeat_region   /rpt_family="Alu"   85895..86028)
repeat_region   /rpt_family="MER4"   complement(86180..86477)
repeat_region   /rpt_family="Alu"   86497..86780
repeat_region   /rpt_family="Alu"   87125..87256
repeat_region   /rpt_family="Alu"   87590..87893
repeat_region   /rpt_family="Alu"   89103..89401
repeat_region   /rpt_family="MER4"   complement(89726..89837)
repeat_region   /rpt_family="MER4"   complement(89888..90390)
repeat_region   /rpt_family="MER4"   complement(91891..92197)
repeat_region   /rpt_family="LTR12"   complement(93951..94017)

BASE COUNT   23552 a   24227 c   24425 g   22834 t
ORIGIN

Alignment Scores:
Pred. No.:   2.37e-188
Score:   2622.50
Percent Similarity:   39.20%
Best Local Similarity:   81.49%
Query Match:   9

Length:   95038
Matches:   624
Conservative:   0
Mismatches:   1
Indels:   968
Gaps:   4

US-10-029-347-2 (1-625) x AF015416 (1-95038)

Qy   1   MetLeuAlaGlnProGlnLeuArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla   20
Db   65026   ATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCCGGGACGCTGCCCGCG   65085

Qy   21   LeuGlyGlyGlyArgProGluAlaAlaProCysThrAspSerProPheGluAlaSerGlyAlaArg   40
Db   65086   CTGGGGGGCGGCCGCCCTGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCAGCGGCGCCGGG   65145

Qy   41   ValLeuGlyGlyLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr   60
Db   65146   GTGCTAGGCGGGGCTGCTCGAGCAAGGCGCTGCTGCCCACCGCCCTCCTGCTGGTGACCACG   65205

Qy   61   ArgAlaAlaAlaProGlyArgLeuGlnGlyLeuArgLeuCysSerProGlnCysAlaGluVal   80
Db   65206   CGCGCCGCCGCGCCCCGGAGGCTGCAGGGCCTGCGCCTGTGTTCCCCGCAGTGCGCCGAGGTG   65265

81   ArgGlyPheSerAspLysLysTyrPheTyrLysTyrPheArgAspGluArg   100
Qy
Db   65266   CGGGGCTTCTCCGACAAGAAGTACTTCTACAAGTACTTCCGGGATGAGAGG   65325

101   ArgAlaArgAlaTyrArgPheLysGluAsnGluThrLeuPheAlaLeuCysPhe   120
Qy
Db   65326   AGGGCCAGGGCCTACCGGTTCAAGGAGAACGAGACGCTGTTCGCCCTGTGCTTC   65385

121   ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGluGlyAlaGlyTyrArg   140
Qy
Db   65386   GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGGCAGCAGCTGGAGGGCGGCTCCGGCGG   65445

141   AspLeuSerArgThrSerLysThrThrThrTyrLeuPheThrSerVal   160
Qy
Db   65446   GACCTGTCGCGCACGTCCAAGACCACCACGTACCTGCTTTCATCACCAGCGTT   65505

161   LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyLeuAspArgLeuArgAsnLeuCys   180
Qy
Db   65506   CTGAGTCGGCTCCGGTACCGGCCCGTTGCAGGCGCACCTGCGACCTGCTGCAATCTGCGC   65565

181   ArgLeuAlaAlaArgGluGlyLeuArgArgAlaGlnPheAlaAlaGlnLysGluGluGlu   200
Qy
Db   65566   CGCCTGGCCGCCGGGGAGGGCCTCCGGACGCAGGGCCAGTTGCCGAGAAGGAACTGGAG   65625

201   GlnLeuGluLeuArgGlySerLysSerValGlnThrLeuPheLeuSerLysLysGluLeuPro   220
Qy
Db   65626   CAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCCGTGTTTCTCAGCAAAAAGGAGCTGCCG   65685

221   GlyValLeuGluThrThrLysGlnPheIleAspGlnSerPheGlnSerPheLeu   240
Qy
Db   65686   GGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCAGAGCTTCCAGAGTTCCTC   65745

241   AlaAlaLeuSerTyrLeuGluLeuGluAspGlyGlyValProArgThrAlaAlaGlyValGlyVal   260
Qy
Db   65746   GCGGCACTGTCCTACCTGTGGAGACGGGGCGTGCCCAGGACCGCGGCTGGCGGGCTGGGCGT   65805

261   GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrHisThrArgPhe   280
Qy
Db   65806   GGGACACTCCTGCGTGGGGACGCCCAGCCACTTGGTGCTCACCACGCGCTTC   65865

281   LeuPheGlyLeuLeuSerArgAlaGluArgMetArgAspIleGluArgHisProGlyCysMet   300
Qy
Db   65866   CTCTTCGGACTGCTGAGCCGCGAGCGGATGCGCGACATCGAGCGCCACTTCGGCTGCATG   65925

301   ValSerGluArgValLysGlnGluValThrGluThrGlyLeuAlaLysThrGluAspThrGluGluPro   320
Qy
Db   65926   GTTTCAGAGCGTGTGAAGCAAGAGGCCCTGCGGTGGCTGGCCGACAGGGACAGGGCCTGC   65985

321   ProGlyValAlaProGluValThrGluLeuGlyLeuAlaLysThrGluAspThrGluPro   340
Qy
Db   65986   CCCGGAGTGGCACCAGAGGTGACCGAGCTGGGCCTGGCCAAAGGCTCGAGGACACCGAAGCCA   66045

341   GluGluGluGluGlyGlyLeuProAsnTyrProLeuLeuLeuTyrCysValArgLeuTyrLeuTyr   360
Qy
Db   66046   GAGGAGGAGGAGGGGGGCCTGCCCAACTACCCACTGGAGTTGTGTACGTGCCTGTAC   66105

361   GluThrGlnGluAspAlaPheValAlaArgGlnAlaLeuCysArgProGluLeuAlaLeu   380
Qy
Db   66106   GAGACGCAGGAGGACGCGTTTGTGGCCAGACCCTGTGCCGCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCCCTG   66165

381   GlnArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysCys   400
Qy
Db   66166   CAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGAGGTGCTGC   66225

401   ProAlaGlyGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnGluLysLys   420
Qy
Db   66226   CCTGCTGGCCAGGCACTGCGCCTGATCAGCTGCAGATTGGTTGCTGCCGCAGGAGAAAG   66285

421   LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGlyGly   435
Qy
Db   66286   AAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGAGGTG   66345

```
Qy    435 --------------------------------------------------                                                      435
Db  66346 GCAGAGATACTTCTTGTGTGTGAGGTGTGTGCCGGTGCAAACCTGTGCCACCTCAGTG   Db  67426 GCAGGCTCTGCCCACCAGACGAGAGGAGGGGGATGAGGGTCTGGGGCTTCATTACC   67485
Qy    435 --------------------------------------------------                                                      435
Db  66406 TCCATCTGGCATGGGGCGCACCTCCAGACCCTGGCTCCCCAGATGGGGAGGCTTT   66465   Db  67486 AGGTTAGGTACTGGTAGGAGAAAAGGTTGGTCCTGGTGAGTGGAAAGCAGCTAAGCACACTG   67545
Qy    435 --------------------------------------------------                                                      435
Db  66466 GACCACCTCCTTGCAAGGCTACCACTGAATTCCTATCTTCTTCCCAGGTGTCCTCACC   66525   Db  67546 AGGGAGGGGATTAAAGAGACAGAGAGTGTATTTCCTTCTGAAGGGAGGCTGCCGTTGGAA   67605
Qy    435 --------------------------------------------------                                                      435
Db  66526 ATGCCCACCGTCAGCCACCCACCCAACCTTCCTAGTAATGGCCCTACAGTCCTG   66585   Db  67606 ACTGCTGGGTTCTCAGAAGCAGTGGAGGAAGGTGTTCAAGCGCAGGCTCCAAGGTCAGGC   67665
Qy    435 --------------------------------------------------                                                      435
Db  66586 GCCCAGCCCTTCTTCCCATTTTACTTTCTGACAATAAGCTCTGGAACACAGATGCCCTC   66645   Db  67666 TGTCTGCAAATGCTAGCCGCCTTCTGCGTGCCCAGTAGCAAGTCCATCTGTCCGAGCTTC   67725
Qy    435 --------------------------------------------------                                                      435
Db  66646 TGTCCCAGAGCCTCTGCACTTGCCCTTCCCTCACCTAGAATATTCTGTTTGTGCCTGTC   66705   Db  67726 CATTTCCTCCCCTGTGAAATGGAGCCAGTAGATCCAGAGAGTTGGGGTGAGATTTAATGT   67785
Qy    435 --------------------------------------------------                                                      435
Db  66706 CTGCAAGTCTGCCCTGGCTCCCCCAGCTGGCTGGGGACCCACTGGCTCCAGAAGTACCC   66765   Db  67786 GGCTACACATGTACAGTCTTTTTTTTTTTTTTTTTGAGACCGAGCCTCGCTCTGTC   67845
Qy    435 --------------------------------------------------                                                      435
Db  66766 TTTGACCACCCTGCTGGACATCTGCCCCACCTCCACCAGACCCTCGGCTCCTCCAG   66825   Db  67846 GCCCAGGCTGGAGTGCAGTGGTGTGCATCTCAGCTCACTGCAAGTCTCCGCCTCCCGGGTTC   67905
Qy    435 --------------------------------------------------                                                      435
Db  66826 GGCAAGGCTTTGCCAGAGTTCCTTGAATATAGCTCCAATTCAGCCCCTGCCCAGGGGC   66885   Db  67906 ACGCCCATTCTCCTGCCTCAGCCTCCCGAGTAGCTGGGACTACAGGTGCCCGCCACTACGC   67965
Qy    435 --------------------------------------------------                                                      435
Db  66886 TTGGCATCTGTTGAGAGGATGAATGTCACGGTGCTGGCAGGGGCTGGGGCTGGGGACCCA   66945   Db  67966 CCGGCTAATTGTTTGTATTTTTAGTAGAGACAGGGTTTCACCGGGTTAGCCAGGATAGTC   68025
Qy    435 --------------------------------------------------                                                      435
Db  66946 AAGACCCCATGGTAGGGGATGTGGGTGATACCGTCCCAGTGACCTCACTCAGACACCCAGT   67005   Db  68026 TCGATCTCCTGACCTCGTGATCCACCCGCCTCGCCTCCCAAAGTGCTGGGATCACAGGC   68085
Qy    435 --------------------------------------------------                                                      435
Db  67006 GGGATGATGCCTGGGCCACGATGAGTGGTTTTGCTCTGGGTTGGTTGGGGGAGTGTGA   67065   Db  68086 ATGAGGCACCGCCCGCCCACATGTACAGTTCTTACATGCAATAACTTGCCAGTCACT   68145
Qy    435 --------------------------------------------------                                                      435
Db  67066 AGGGGGTGGCTCAGCAAGGAGTGAGGAGGGGCAGCTCTGACCCCAGGACTACAG   67125   Db  68146 CTTAGCTACCACCAAGTGAGCAGAGAAGGAAGAGAGAAATCTAGGAACTAGTACTGATGGTG   68205
Qy    435 --------------------------------------------------                                                      435
Db  67126 ATAGACAGGAGTCCTTTCCGGCAGGCACTGTGAGAGGTGGGCTTCACCTCCTCT   67185   Db  68206 AAACAGGTCCATCCTCCTTACGGTGATGCAACTCGGCAAATTAGCTTGATATGGGGGTGGGGG   68265
Qy    436 --------SerSerGlnGlyThrThrLysGlnLeuProAlaSerLeuHisProLe  452   Qy    467 --------------------------------------------------                                                      467
Db  67186 CTGTCCCAGCAGTTCTCAAGGACCACCAAAAACTGCCAGCCTCCCTTCTTCATCCACT   67245   Db  68266 ATTTGCAATCTGAGCATAAGAACCAAAAAGGTGATCACATTTGCCCAGCAAT   68325
Qy    452 uPheGlnAlaMetThrAspProLeuCysHisLeuSerSerLeuThr-----  467   Qy    467 --------------------------------------------------                                                      467
Db  67246 CTTTCAGGCAATGACTGACCCCACTGTGCCATCTGACCTTCACGTGAGTGGCCACACC   67305   Db  68326 TCTGCTGGTAAAGTAAGATGTTCAGCACACCATTATTATAATGGAAAAAAAGGA   68385
Qy    467 --------------------------------------------------                                                      467
Db  67306 CCCAGCTCTTCCCACGGAGCGGGCTGAGCAGCCAGGATGCCCGGGCAGAGACGGAAGTATG   67365   Db  68386 GAAACAACACAACTACCCAACCATTAGGAATAGCTGACTAAAATAAACTGAGTTGGCCAC   68445
Qy    467 --------------------------------------------------                                                      467
Db  67366 ACCTAGGAAAGAAGGCTGAGGGGTCAGGGTCAGAGCCTGAGCCTGTGGGGTCTGGAGATG   67425   Db  68446 CTCATCTTGGTGGACGATGTGCGGCAAACATGTTGAAACACGTCTCAGCTCTCCGAAGA   68505
Qy    467 --------------------------------------------------                                                      467
```

```
Qy    579 -------Leu-SerLeuAlaSerValGluLeuGlnSerLeuGlnGluGln 595
Db  69645 GCGTGTGGCTGCAGTCTGGCCTCTGGGAGCTGAGCAGCAGTCACTACAGGAGCTTCAG 69704

Qy    596 AlaValLysArgAlaLysProAspLeuValIleThrHisProAlaLeuAspGlyHisPro 615
Db  69705 GCTGTGAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCCTGGACGGCCACCCA 69764

Qy    616 GlnProProLysGluLeuIleSerThrPhe 625
Db  69765 CAACCTCCCAAGGAACTCATCTCGACCTTC 69794
```

RESULT 16
AC136475

LOCUS       AC136475               143835 bp    DNA     linear   PRI 30-MAY-2002
DEFINITION  Homo sapiens chromosome 11, clone RP11-326C3, complete sequence.
ACCESSION   AC136475
VERSION     AC136475.6  GI:31193978
KEYWORDS    HTG.
SOURCE      Homo sapiens (human)
  ORGANISM  Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE   1  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C. and Lander,E.
  TITLE     Homo sapiens chromosome 11, clone RP11-326C3
  JOURNAL   Unpublished
REFERENCE   2  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Allen,N., Ali,A., Allen,N.,
            Barna,N., Bastien,V., Bloom,T., Boguslavky,L., Boukhgalter,B.,
            Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,
            Cook,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S., Dodge,S.,
            Faro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,
            Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hafez,N.,
            Hagos,B., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., MacLean,C., Macdonald,P., Major,J.,
            Matthews,C., McCarthy,M., Meldrim,J., Meneus,L., Mihova,T.,
            Mlenga,V., Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C.,
            Norman,C.H., O'Connor,T., O'Donnell,P., O'Neil,D., Oliver,J.,
            Peterson,K., Phunkhang,P., Pierre,N., Raymond,C., Retta,R.,
            Rise,C., Rogov,P., Roman,J., Roy,A., Schauer,S., Schupback,R.,
            Seaman,S., Severy,P., Smith,C., Spencer,B., Stange-Thomann,N.,
            Stojanovic,N., Talamas,J., Tesfaye,S., Theodore,J., Topham,K.,
            Travers,M., Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (02-NOV-2002) Whitehead Institute/MIT Center for Genome
            Research, 320 Charles Street, Cambridge, MA 02141, USA
REFERENCE   3  (bases 1 to 143835)
  AUTHORS   Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
            Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
            Boguslavkly,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
            Collymore,A., Cook,A., Cooke,P., Corum,B., DeArellano,K.,
            Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
            Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
            Graham,L., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
            Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
            Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
            Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., MacLean,C.,
            Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
            Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
            Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
            O'Neil,D., Oliver,J., Peterson,K., Phunkharg,P., Pierre,N.,
            Rachupka,A., Ramasamy,U., Raymond,C., Retta,R., Rise,C., Rogov,P.,
            Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
            Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
            Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M.,
            Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
            Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
  TITLE     Direct Submission
  JOURNAL   Submitted (10-APR-2003) Whitehead Institute/MIT Center for Genome
```

```
Db  68506 GCTCTAGAGAGTGTCGCGCAGGGAGAAAGTGAACAGAGCTCCCGGGGTGCATGGTGACAA 68565
Qy    467 ------ 467
Db  68566 AGGGCAGGGCAGGCTCTGCGGGGTCCAGGGAAGGATGTGGCACCAACACATCTCTCAGCT 68625
Qy    467 ------ 467
Db  68626 GAACCCTGGCGCCACCGGGCAGCGGGCATTTGTGCAGTGGAGGTGGTGCGGTGGCTGAG 68685
Qy    468 ------LeuSerHisCysLysLeuProA 475
Db  68686 GCGCCTGCAGGTAAATCTCTGCTTCTTGACCACAGGTGTCCACTGCAAACTCCCTG 66745
Qy    475 spAlaValCysArgAspLeuSerGluAlaArgGluAlaAlaProAlaLeuThrGluLeuG 495
Db  68746 ACGCGGTCTGCCGAGACCTTCTGAGCGCCCGAGGGCAGCCCCCGCACTGAGGGAGCTGG 68805
Qy    495 lyLeuLeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGluyLeuAla 515
Db  68806 GCCTCCTCCACAACAGGCCTCAGTGAGGCAGGACTGCGTATGCTGAGTGAGCGGCCTAGCC 68865
Qy    515 rpProGlnCysArgValGlnThrVal 523
Db  68866 GGCCGCAGTCAGGGTGCAGGACGGT-CAGTGAGGGCCTGGCCTGGGAGGGACCGTGGGAT 68924
Qy    524 ------ArgValGlnLeuProAspP 530
Db  68925 GCCCCGCCACCCAGCAGCTCCTGAGGTCGGCCCTCCCACAGGTACAGCTCCCTGACC 68984
Qy    530 roGlnArgGlyLeuGlnTyrLeuValGlyMetLeuValGlySerProAlaLeuThrThrL 550
Db  68985 CCCAGGCGAGGGGCTCCAGTACCTGGTGGGTATGCTTCGGCAGAGCCCTGACCCTGACCACCC 69044
Qy    550 euAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuG 570
Db  69045 TGGATCTCAGCGGGCTGCCAACTGCCCCCATGTGACCTACCTGTGTGCAGTCCTGC 69104
Qy    570 lnHisGlnGlyCysGlnLeuGlnThr 578
Db  69105 AGCACCAGGATGCGGCCTGCAGACCCTGCAGGTGAGGCAGCAGGGGTGGGAAGGGGTGCTGG 69164
Qy    578 ------ 578
Db  69165 GACACAGCCTGTCAACCCGGCGAGGGGGAGGGTGCCTGGGGGCTTCCCCTGACCCTGAA 69224
Qy    578 ------ 578
Db  69225 GAGGGCCCCCCAGAGACCTCCCATGTGACTGACCTGAGC-CAAACTGACCTGGGAGCCCA 69284
Qy    578 ------ 578
Db  69285 GCCCTGACAGCCCTGCCCCTGAAGCCCTGAGCCCACAGCTCGGGGCCTTGACCACTTGTC 69344
Qy    578 ------ 578
Db  69345 TGACCAAATGTCATCGTGTCTCGTGGCCCATGGCACCCTGGGCCAGTGGCCCCGA 69404
Qy    578 ------ 578
Db  69405 TACCTTTGCCACCCCAAGGTCGGCTCTGCCCCCCCCCCCCAGCCTCTGACAAACCT 69464
Qy    578 ------ 578
Db  69465 CTCAACTGACACCACTGTCCCTGAGTGTCTGTGGCACAGCCCCAGTCA 69524
Qy    578 ------ 578
Db  69525 CTGCCCCGCAGCCGCGGCCGCGCCCCCACACGCACTGCCCGTCACTGCCCGCCGGCGCCCGCTGCCC 69584
Qy    578 ------ 578
Db  69585 CCACGGCACTGCCCCCCAAGCCCTGGCTGCTTCCCCACCGGTGACCCCGCTTCTGTCT 69644
```

```
REFERENCE    4   (bases 1 to 143835)
AUTHORS      Research, 320 Charles Street, Cambridge, MA 02141, USA
             Birren,B., Nusbaum,C., Lander,E., Abouelleil,A., Allen,N.,
             Anderson,S., Arachchi,H.M., Barna,N., Bastien,V., Bloom,T.,
             Boguslavskiy,L., Boukhgalter,B., Camarata,J., Chang,J., Choepel,Y.,
             Collymore,A., Cook,A., Cooke,P., Corum,B., Dearellano,K.,
             Diaz,J.S., Dodge,S., Dooley,K., Dorris,L., Erickson,J., Faro,S.,
             Ferreira,P., FitzGerald,M., Gage,D., Galagan,J., Gardyna,S.,
             Graham,J., Grand-Pierre,N., Hafez,N., Hagopian,D., Hagos,B.,
             Hall,J., Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C.,
             Kamat,A., Karatas,A., Kells,C., Landers,T., Levine,R.,
             Lindblad-Toh,K., Liu,G., Lui,A., Mabbitt,R., Maclean,C.,
             Macdonald,P., Major,J., Manning,J., Matthews,C., McCarthy,M.,
             Meldrim,J., Meneus,L., Mihova,T., Mlenga,V., Murphy,T., Naylor,J.,
             Nguyen,C., Nicol,R., Norbu,C., O'Connor,T., O'Donnell,P.,
             O'Neil,D., Oliver,J., Peterson,K., Phunkhang,P., Pierre,N.,
             Rachupka,A., Ramasamy,J., Raymond,C., Retta,R., Rise,C., Rogov,P.,
             Roman,J., Schauer,S., Schupback,R., Seaman,S., Severy,P., Smith,C.,
             Spencer,B., Stange-Thomann,N., Stojanovic,N., Stubbs,M.,
             Talamas,J., Tesfaye,S., Theodore,J., Topham,K., Travers,M., Wu,X.,
             Vassiliev,H., Venkataraman,V.S., Viel,R., Vo,A., Wilson,B., Wu,X.,
             Wyman,D., Young,G., Zainoun,J., Zembek,L., Zimmer,A. and Zody,M.
TITLE        Direct Submission
JOURNAL      Submitted (30-MAY-2003) Whitehead Institute/MIT Center for Genome
             Research, 320 Charles Street, Cambridge, MA 02141, USA
COMMENT      On May 30, 2003 this sequence version replaced gi:29469599.
             All repeats were identified using RepeatMasker:
             Smit, A.F.A. & Green, P. (1996-1997)
             http://ftp.genome.washington.edu/RM/RepeatMasker.html
             --------- Genome Center
             Center: Whitehead Institute/ MIT Center for Genome Research
             Center code: WIBR
             Web site: http://www-seq.wi.mit.edu
             Contact: sequence_submissions@genome.wi.mit.edu
             ---------
             Center project name: L28535
             Center clone name: 326_C_3
             ---------
             Only the first 143.8 kilobases of this clone are being submitted.
             The remainder overlaps accession number AC138235 [WICGR project
             L29013].
FEATURES             Location/Qualifiers
     source          1..143835
                     /organism="Homo sapiens"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:9606"
                     /chromosome="11"
                     /map="11"
                     /clone="RP11-326C3"
                     /clone_lib="RPCI-11 Human Male BAC"
     repeat_region   1712..2006
                     /rpt_family="AluSg"
     repeat_region   2347..2703
                     /rpt_family="THE1B"
     repeat_region   4157..4302
                     /rpt_family="L1ME4a"
     repeat_region   4435..4455
                     /rpt_family="AT_rich"
     repeat_region   4528..4591
                     /rpt_family="AT_rich"
     repeat_region   4592..4901
                     /rpt_family="AluSx"
     repeat_region   4902..4968
                     /rpt_family="AT_rich"
     repeat_region   4979..5091
                     /rpt_family="L1ME4a"
     repeat_region   5154..5251
                     /rpt_family="L1ME4a"
     repeat_region   5620..5683
                     /rpt_family="L1ME4a"
     repeat_region   complement(5684..5954)
                     /rpt_family="AluJo"
     repeat_region   5955..6050
                     /rpt_family="L1ME4a"
     repeat_region   complement(6059..6206)
                     /rpt_family="MER46C"
     repeat_region   6349..6578
                     /rpt_family="L1ME4a"
     repeat_region   6792..7203
                     /rpt_family="L2"
     repeat_region   complement(8884..9052)
                     /rpt_family="L1MC5"
     repeat_region   9225..9287
                     /rpt_family="MER68-int"
     repeat_region   9676..9986
                     /rpt_family="AluSg"
     repeat_region   9994..10032
                     /rpt_family="AT_rich"
     repeat_region   10182..10417
                     /rpt_family="L1MEd"
     repeat_region   complement(10418..10946)
                     /rpt_family="LTR59"
     repeat_region   complement(10947..11251)
                     /rpt_family="AluSx"
     repeat_region   complement(11252..11272)
                     /rpt_family="AluSx"
     repeat_region   complement(11290..11595)
                     /rpt_family="LTR59"
     repeat_region   11702..12014
                     /rpt_family="AluSp"
     repeat_region   12047..12137
                     /rpt_family="AluSg"
     repeat_region   complement(12159..12465)
                     /rpt_family="L1M4"
     repeat_region   12465..12605
                     /rpt_family="AluSx"
     repeat_region   12763..12832
                     /rpt_family="L2"
     repeat_region   12898..13715
                     /rpt_family="L2"
     repeat_region   13716..14007
                     /rpt_family="L1MB5"
     repeat_region   14018..14320
                     /rpt_family="AluSc"
     repeat_region   14321..14603
                     /rpt_family="AluSx"
     repeat_region   4611..14772
                     /rpt_family="L1MB5"
     repeat_region   15386..15685
                     /rpt_family="AluSg/x"
     repeat_region   15701..15984
                     /rpt_family="AluSx"
     repeat_region   complement(16654..16949)
                     /rpt_family="AluJo"
     repeat_region   17027..17095
                     /rpt_family="AluSg"
     repeat_region   complement(18948..19258)
                     /rpt_family="L1MC5"
     repeat_region   19299..19409
                     /rpt_family="AluY"
     repeat_region   complement(19426..19719)
                     /rpt_family="L2"
     repeat_region   complement(19742..20040)
                     /rpt_family="AluSg"
     repeat_region   20207..20435
                     /rpt_family="AluJo"
     repeat_region   20906..21032
                     /rpt_family="L1MC/D"
     repeat_region   complement(21181..21341)
                     /rpt_family="(CCCCGG)n"
     repeat_region   complement(22663..22828)
                     /rpt_family="MIR"
```

Alignment Scores:
Pred. No.:          143835
Score:              6.83e-188    Length:
                    2619.50      Matches:    623

us-10-029-347-2.rge

```
Percent Similarity:    39.20%      Conservative:    .
Best Local Similarity: 39.13%      Mismatches:      1
Query Match:                       Indels:          969
DB:                                Gaps:            4

US-1C-029-347-2 (2-625) x AC136475 (1-143835)
```

```
Qy      1 MetLeuAlaGlnProGlnArgLeuLeuIleLeuAspGlyAlaAspGluLeuProAla  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65077 ATGCTGGCGCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGGGCGGACGAGCTGCCGGCG  65136

Qy     21 LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65137 CTGGGGGGCCCCGAGGCCGCCCCTTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCGGG  65196

Qy     41 ValLeuGlyGlyLeuLeuSerLeuAlaLeuLeuProThrAlaLeuLeuLeuValThrThr  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65197 GTGCTAGGCGGCGGGCTGCTGAGCAAGGCGCTGCTGCCACGGGCCCTCGTGGTGGACCACG  65256

Qy     61 ArgAlaAlaAlaProGlyArgLeuGlnCysLeuCysSerProGlnCysAlaGluVal  80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65257 CGGCCGCCGCCCCCGGGAGGCTGCAGGGCGCCGCCTGTTCCCCGCAGTGCGCCGAGGTG  65316

Qy     81 ArgGlyPheSerAspLysAspLysLysLysTyrPheTyrLysPhePheAspPheArgGluArg  100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65317 CGGGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTATTCCGGGATGAGAGAG  65376

Qy    101 ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65377 AGGGGCCAGGCGCGCCTACCGGTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC  65436

Qy    121 ValProPheValCysTrpIleValCysThrValArgGlnArgGluLeuGlyLeuGlyArg  140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65437 GTGCCCTTCGTGTGCTGGAT:GTGTGCACCGTGCGCCAGCGTGGAGCTGGGCTCGGCGG  65496

Qy    141 AspLeuSerArgThrSerLysThrThrThrSerValTyrLeuLeuPheIleThrSerVal  160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65497 GACCTGTCGCGCACCTCCAAGACCACCACCAGCGTCAGTGTACCCGCTTTCATCACCAGCGT  65556

Qy    161 LeuSerSerAlaProValAlaAspGlyHisGlyHisAspLeuArgLeuAsnLeuCys  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65557 CTGAGCTCGGCTCCGGTAGCGGACGGGCACCGCCGCCCAATCGTGTGC  65616

Qy    181 ArgLeuAlaArgGluGlyValLeuGlyValGlyArgGlnPheAlaGlnGluLysGluLeuGlu  200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65617 CGGCTGGCGCGGGAGGGCGTGCTCGGCGTGGGACGGCAGTTTGCCCAGGAGAAGGAACTGGAG  65676

Qy    201 GlnLeuGluArgGlySerLysSerValGlnThrLeuPheLeuSerLysLysGluLeuPro  220
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65677 CAACTGGAGCTTCG:GGCTCAAGTGCAGACGCTGTTTCTCAGCAAAAAGAAGAGCTGCCG  65736

Qy    221 GlyValLeuGluThrGluValThrTyrGlnPheIleAspPheSerPheGluProLeu  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65737 GGCGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACCACCAGAGCTTCCTGGAGTTCCTC  65796

Qy    241 AlaAlaLeuSerTyrLeuGluGluAspProArgThrAlaAlaAlaGlyYGlyVal  260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65797 GCGGCACTGTCCTACCTGCTGGAGGAGGACCCGCGGACCGCGGCTGGGCGGGCTTC  65856

Qy    261 GlyThrLeuLeuGlyAspAlaGlnLnProHisSerHisLeuValLeuThrThrArgPhe  280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65857 GGGACACTCCTCGGCGGGGAAGACCCCCAGCCACTGTGCTCACCACGCGCTTC  65916

Qy    281 LeuPheGlyLeuLeuSerAlaGluArgMetArgAspIlePheGlyGlyGlyGlyCysMet  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65917 CTC-TCGGACTGTCGGCGGAGCGGATGCGCGACATCGAGGGCGGGCACTCGGCTGCATG  65976

Qy    301 ValSerGluArgValLysGlyGlnGluLeuAlaLeuAlaLeuArgTrpValGlnGlyGlyGlyCys  320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  65977 GTTTCAGAGCGTGTGAAGGCAGCAGAGCCCTGCGGTGGGTGCAGGGGCTGCAGGGACAGGGCTGC  66036

Qy    321 ProGlyValAlaProGluVal:ThrGluGluGluGlyLeuLysGlyAlaLysValAspThrGlu  340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66037 CCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGGCTCGAGGACACCGAAGAGCCA  66096

Qy    341 GluGluGluGlnGluGlyLyGluGluProAsnTyrProLeuLeuLeuGlnTyrCysLeuTyr  360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66097 GAGGAGGAGGAGGAGGAGGAGCCCAACTACCCACTGGAGTTGCTGCTACTGCCTGTAC  66156

Qy    361 GluThrGlnGluGluArgProAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu  380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66157 GAGACGCAGGAGGAGCGCCCGGCCTTTGTGCGCCAAGCTGCCTGGGCGGTGGGCGCTG  66216

Qy    381 GlnArgValArgPheCysArgMetAspProValAlaLeuValLeuSerTyrCysValArgCysCys  400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66217 CAGCGAGTGCGGCTTCTGCCGCATGGACCCGGTGTTCGTGTCGTGAGTGTTGCGTGAGTGTCTGC  66276

Qy    401 ProAlaGlyGlnAlaLeuAlaLeueuSerCysArgGleuValAlaAlaGlnGluLysLys  420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  66277 CCTGCTGGACAGCGACTGCCGGCTGATCAGTTGCGCAGATTGGTTGCTGCGCAGGAAGAAG  66336

Qy    421 LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGly-----------  435
          ||||||||||||||||||||||||||||||||
Db  66337 AAGAAGAGCCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGCGGCGGCAGTGGTGGTCTCCAGG  66396

Qy    435 -----------------------------------------
Db  66397 GCAGAGATACTCTCTTGTGTGGAGGTGTGTGCCGGTGCAAACCTGTGCACCTGTCAGTG  66456

Qy    435 -----------------------------------------
Db  66457 TCCATCTGGCAATGGGCGCACCTGCCTCCCAGACCCTGGCTCCCAGATGGGAGGAGCTTT  66516

Qy    435 -----------------------------------------
Db  66517 GACCACCTCCTTGCAAGGCAACCACTGAATTCCTATCTTCTTCCCCAGGGTGTCCTCACC  66576

Qy    435 -----------------------------------------
Db  66577 ATGCCCACCGTCAGCCACCACCCACAACCTATGAAAACTTCTAGTATGGCCCTACAGCTCTG  66636

Qy    435 -----------------------------------------
Db  66637 GCCCAGCCCCTTCTTCCCATTTACTTCTGACAATAGCTCTGGAACACAGATGCCCTC  66696

Qy    435 -----------------------------------------
Db  66697 TGTCCAGAGCCTCTGCACTTGCCCTTCCTTCCTCCACCTAGAATATTCTGTTTGTGCCTGTC  66756

Qy    435 -----------------------------------------
Db  66757 CTGCAAGTCTGCCCTGGCTCCCCCAGTGGCTGGGGACCCACCTGGCTCCAGAAGTACCC  66816

Qy    435 -----------------------------------------
Db  66817 TTTCGACCACCCTGCTGTGACATCTGCCCCCACCTCCACCAGACCCTCGGCTCCTCCAG  66876

Qy    435 -----------------------------------------
Db  66877 GGCAAGGCTTTGCCAGAGTTCCTTGAATATAGCTCCAATTCAGCCCCTGCCCAGGGGC  66936

Qy    435 -----------------------------------------
Db  66937 TTGGCATCTGTTGAGAGATGAATGTCACGGTGTTGTTGCCAGGGCCTGGGCCTGGGGACCCA  66996

Qy    435 -----------------------------------------
Db  66997 AAGACCCATGTAGGAGGATGTGGGTGATACCGTCCCACTGACCTCACTCAGAACACCCAGT  67056

Qy    435 -----------------------------------------
Db  67057 GGGATGATGCCTGGGCCCACAGATAGGTGTTTGCTCTGGGTTGGGTTGGGGAGTGTGA  67116

Qy    435 -----------------------------------------
Db  67117 AGGGGGTGGCTCAGCAGGAGGTGAGGAGGAGGGGCAGCTCTGAGCACCCCAGGACTACAG  67176
```

```
Qy     435 ------------------------------------------------------------  435
Db   67177 ATAGACAGGAGTCCTTCCGGCACTGTTGATGGCAACTGGCAAATTAGCTTGATATGGGGTGGGGG  68316
                                                                          467

Qy     436 ------SerSerGlnGlyThrThrLysGlnLeuProAlaSerLeuLeuHisProLe        452
                 |||||||||||||||||||||||||||||||||||||||||||
Db   67237 CTCCCCAGCAGTTCTCAAGGACCACAAAACAACTGCCAGCTCCCTTCTTCATCCACT  68376
                                                                          467

Qy     452 uPheGlnAlaMetThrAspProLeuCysHisSerLeuThr------------------        467
           ||||||||||||||||||||||||||||||||||
Db   67297 CTTTCAGGCAATGACTGACCCACTGCCACTGTGCCATCTGAGCAGCCTCACGTGAGTGCCACACC  68436
                                                                          467

Qy     467 ------------------------------------------------------------  467
Db   67357 CCCAGCTCTTCCCACGGAGCGGGCTGAGAGCCTGGGCTGAGAGCAGGAGAGGAACGGAAGTATG  68496
                                                                          467

Qy     467 ------------------------------------------------------------  467
Db   67417 ACCTAGGAAAGAGAGGCTCAGAGACCCAGCCCTGAGCGCTGTGGGTGCTGGAGATG  68556
                                                                          467

Qy     467 ------------------------------------------------------------  467
Db   67477 GCAGGGCTGCCCACACAGAGCAGAGGGAGGAGGGGATGAGGGTCTTGGGGCTTCATTACC  68616
                                                                          467

Qy     467 ------------------------------------------------------------  467
Db   67537 AGGTTAGGTACTGGTAGGAAAAGGTTGGTCCTGGTGAGTGGAAAGCAGCTAAGCACACTG  68676
                                                                          467

Qy     467 ------------------------------------------------------------  467
Db   67597 AGGGAGGGGATTAAAGAAGACAAGAGTGTATTTCCTTCTGAAGGGAGGCTGCCGTGTGGAA  68736
                                                                          467

Qy     467 --------------LeuSerHisCysLysLeuProA        475
                         ||||||||||||||||||||
Db   67657 ACTGCTGGGTTCTCAGAGACCAGTGGAGGAAGTGTTCAAGCCGCTCCAGGTCCAGGTCAGGC  68796
                                                                          467

Qy     475 spAlaValCysArgAspLeuSerGluAlaLeuArgAlaArgAlaAlaProAlaLeuThrGluLeuG  495
           |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67717 ACGCGGGTCTGCCCACGTCCCAGTAGCAAGTCCATCTGTCCGAGCTTC  68856
                                                                          467

Qy     495 lyLeuLeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuLeuAlaT  515
           ||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67777 GCCTCCTCCACACAGCTCAGTGAGCGGGGACTGCGTATGCTGAGTGAGGGCCTAGCCC  68916
                                                                          467

Qy     515 rpProGlnCysArgValGlnThrVal----------------        523
           |||||||||||||||||||||||||||
Db   67837 GGCCGCAGTGAGGGGTGCAGACGGT-CAGTGAGGCCTGGCCTGGAGGGGACCGTGGGAT  68975
                                                                          467

Qy     524 -------------------------ArgValGlnLeuProAspP        530
                                    ||||||||||||||||||||
Db   67897 GCCCAGGCTGGAGTGTGAACTCAGTGAACTCACTGCAAGCTCCCGGGTTC  69035
                                                                          467

Qy     530 roGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSerProAlaLeuThrThrL  550
           ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   67957 CCCAGGCGGGGCTCCAGTACCTGGTGGGTATGCTTCGGCAGCCCCGCCCTGACCACCC  69095
                                                                          467

Qy     550 euAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuG  570
           ||||||||||||||||||||||||||||||||||||||||||||
Db   68017 TGGATCTCAGCGGCTGCCAACTGCCCCATGGTGACCTACCTGTGTGCAGTTCTGC  69155
                                                                          467

Qy     570 lnHisGlnGlyCysGlyLeuGlnThr-        578
           ||||||||||||||||||||||||
Db   68077 AGCACCAGGGATGCGGCCTGCAGACCCTCAGGTGGAGGGGTGGGAAGGGGTGCTGGGG  69215
                                                                          578

Qy     578 ------------------------------------------------------------  578
Db   68137 ATGAGCCACCGCGCCGCCGCCAATGTACAGTTCTTACATGGCAATAACTTGCCAGTCACT  69275
                                                                          578

Qy     578 ------------------------------------------------------------  578
Db   68197 CTTAGCTACCACCAAGTGAGCAGAAGGAAGAGAGAAATCTAGGAACTAGTACTAGTGATGGTG  69335
                                                                          578
```

Db    69336  GCCCTGACAGCCTGCCCTCTGAAGCCCTGGCCACAGTCGGGGCCTTGACCACTTGTC 69395

Qy      578  ------------------------------------------------------------  578

Db    69396  TGACCAAATGTTCATCTGTTGTCTGGTGCCCATGGCAGCACCCCTGGGCCAGTGCCCCGA 69455

Qy      578  ------------------------------------------------------------  578

Db    69456  TACCCTTTGGCACCCCAAGTCCGGCCCTCTGCCCCTCCCCGAGCCTCTGACAAACCT 69515

Qy      578  ------------------------------------------------------------  578

Db    69516  CTCAACTGACACCACCTGTCCCTGAGTGTCTGTGGCACAGCCAAGCTGACAGCCCAGTCA 69575

Qy      578  ------------------------------------------------------------  578

Db    69576  Leu-SerLeuAlaSerValGluLeuGly-GlnSerLeuGlnGluGluGln 595
             -----------------|||||||||||||||||||||||||||||||||
             CTGCCCGCAGCCCGGCCACCCCACGGCACTGCCCGTCACTGCCGGCCCCGGCCCGGCTGCCC 69635

Qy      578  ------------------------------------------------------------  578

Db    69636  CCACGGCACTGCCCCCAAGCCTGGCTGCTTCCCCACCGCTGACCCCGGCTTCTGCTCT 69695

Qy      579  Leu-SerLeuAlaSerValGluLeuGly-GlnSerLeuGlnGluGluGln 595
             ---|||||||||||||||||||||||||||||||||||||||||
Db    69696  GCGTGTGGCTGCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAG 69755

Qy      596  AlaValLysArgAlaLysProAspLeuValIleThrHisProAlaLeuAspGlyHisPro 615
             ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    69756  GCTGTGAAGAGAGCAAAGCCGGATCTGGTCATCACACACCCAGCCTGGACGACGCCACCCA 69815

Qy      616  GlnProProLysGluLeuLeuSerThrPhe 625
             ||||||||||||||||||||||||||
Db    69816  CAACCTCCCAAGGAACTCATCTCGACCTTC 69845

RESULT 11
AP006283
LOCUS         AP006283           156819 bp    DNA      linear    HTG 26-MAR-2003
DEFINITION    Homo sapiens chromosome 11 clone RP11-326C3 map 11p, WORKING DRAFT
              SEQUENCE, 19 unordered pieces.
ACCESSION     AP006283
VERSION       AP006283.1    GI:29243343
KEYWORDS      HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE        Homo sapiens (human)
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1
AUTHORS       Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE         Homo sapiens genomic DNA of 11p
JOURNAL       Published Only in Database (2003)
REFERENCE     2  (bases 1 to 156819)
AUTHORS       Hattori,M., Ishii,K., Toyoda,A., Taylor,T.D., Hong-Seog,P.,
              Fujiyama,A., Yada,T., Totoki,Y., Watanabe,H. and Sakaki,Y.
TITLE         Direct Submission
JOURNAL       Submitted (24-MAR-2003) Masahira Hattori, The Institute of Physical
              and Chemical Research (RIKEN), Genomic Sciences Center (GSC);
              1-7-22 Suehiro-chou, Tsurumi-ku, Yokohama, Kanagawa 230-0045, Japan
              (E-mail:hattori@gsc.riken.go.jp, URL:http://hgp.gsc.riken.go.jp/,
              Tel:81-45-503-9111, Fax:81-45-503-9170)
COMMENT       -------------------- Genome Center
              Center: RIKEN Genomic Sciences Center(GSC)
              Center code: RIKEN
              Web site: http://hgp.gsc.riken.go.jp/
              Contact: hattori@gsc.riken.go.jp
              -------------------- Project Information
              Center project name: HumDraft11
              Center clone name: RP11-326C3
              -------------------- Summary Statistics
              Sequencing vector: PCR products; 100% of reads
              Chemistry: Dye-terminator ET-amersham; 100% of reads
              Assembly program: Phrap; version 0.990329

Consensus quality: 144706 bases at least Q40
Consensus quality: 150967 bases at least Q30
Consensus quality: 153657 bases at least Q20
Insert size: 15019; sum-of-contigs
Quality coverage: 6.36x in Q20 bases; sum-of-contigs

----------------
NOTE: This is a 'working draft' sequence. It currently consists of
19 contigs. The true order of the pieces is not known and their
order in this sequence record is arbitrary. Gaps between the
contigs are represented as runs N, but the exact sizes of the gaps
are unknown. This record will be updated with the finished sequence
as soon as it is available and the accession number will be
preserved.

            1  contig of  21001 bp in length
        21102  contig of  19613 bp in length
        40714  contig of  17616 bp in length
        40815  contig of   8224 bp in length
        58531  contig of  14280 bp in length
        66754  contig of   8529 bp in length
        66855  contig of  12026 bp in length
        81235  contig of  10636 bp in length
        89763  contig of   7077 bp in length
       101889  contig of   5279 bp in length
       101990  contig of   3903 bp in length
       112625  contig of   7976 bp in length
       112726  contig of   6182 bp in length
       119802  contig of   3249 bp in length
       119903  contig of   2577 bp in length
       125181  contig of   3010 bp in length
       125282  contig of   1357 bp in length
       129184  contig of   1427 bp in length
       129285  contig of   1057 bp in length
       137260
       137361
       143542
       143643
       146891
       149568
       149669
       152678
       152779
       154135
       154236
       155662
       155763
       156819

* NOTE: This is a 'working draft' sequence. It currently
* consists of 19 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.

*        1:  contig of  21001 bp in length
*    21002:  gap of    100 bp
*    21102:  contig of  19613 bp in length
*    40714:  gap of    100 bp
*    40715:  contig of  17616 bp in length
*    40815:  gap of    100 bp
*    58430:  contig of   8224 bp in length
*    58531:  gap of    100 bp
*    66754:  contig of  14280 bp in length
*    66854:  gap of    100 bp
*    66855:  contig of   8529 bp in length
*    81134:  gap of    100 bp
*    81235:  contig of  12026 bp in length
*    81234:  gap of    100 bp
*    89763:  contig of  10636 bp in length
*    89864:  gap of    100 bp
*    89864:  contig of  12026 bp in length
*   101889:  gap of    100 bp
*   101890:  contig of  10636 bp in length
*   101990:  gap of    100 bp
*   112625:  contig of   7077 bp in length
*   112626:  gap of    100 bp
*   112726:  contig of   7976 bp in length
*   119802:  gap of    100 bp
*   119803:  contig of   6182 bp in length
*   119902:  gap of    100 bp
*   119903:  contig of   3249 bp in length
*   125181:  gap of    100 bp
*   125182:  contig of   5279 bp in length
*   125281:  gap of    100 bp
*   125282:  contig of   2577 bp in length
*   129184:  gap of    100 bp
*   129185:  contig of   3903 bp in length
*   129284:  gap of    100 bp
*   129285:  contig of   1357 bp in length
*   137260:  gap of    100 bp
*   137261:  contig of   7976 bp in length
*   137360:  gap of    100 bp
*   137361:  contig of   6182 bp in length
*   143542:  gap of    100 bp
*   143543:  contig of   3249 bp in length
*   143642:  gap of    100 bp
*   143643:  contig of   3249 bp in length
*   146891:  gap of    100 bp
*   146892:  contig of   3249 bp in length
*   146991:  gap of    100 bp
*   146992:  contig of   2577 bp in length
*   149568:  gap of    100 bp
*   149569:  contig of   3010 bp in length
*   149668:  gap of    100 bp
*   149669:  contig of   3010 bp in length
*   152678:  gap of    100 bp
*   152679:  contig of   1357 bp in length
*   152778:  gap of    100 bp
*   152779:  contig of   1357 bp in length
*   154135:

```
  *   :54136   :54235: gap of 100 bp
  *   :54236   155662: contig of 1427 bp in length
  *   155663   155762: gap of 100 bp
  *   155763   1568:9: contig of 1057 bp in length.

FEATURES       Location/Qualifiers
source         1..156819
               /organism="Homo sapiens"
               /mol_type="genomic DNA"
               /db_xref="taxon:9606"
               /chromosome="11"
               /map="11p"
               /clone="RP11-326C3"
misc_feature   1..21001
               /note="assembly_fragment"
misc_feature   2102..40714
               /note="assembly_fragment"
misc_feature   40815..58430
               /note="assembly_fragment"
misc_feature   58531..66754
               /note="assembly_fragment clone_end:SP6 vector_side:left"
misc_feature   6685..8134
               /note="assembly_fragment"
misc_feature   81235..89761
               /note="assembly_fragment"
misc_feature   89864..101899
               /note="assembly_fragment"
misc_feature   101990..112625
               /note="assembly_fragment"
misc_feature   11726..119802
               /note="assembly_fragment"
misc_feature   11990..125-81
               /note="assembly_fragment"
misc_feature   :25282..129-84
               /note="assembly_fragment clone_end:T7 vector_side:right"
misc_feature   129285..137260
               /note="assembly_fragment"
misc_feature   13736:..143542
               /note="assembly_fragment"
misc_feature   143643..146891
               /note="assembly_fragment"
misc_feature   146992..149568
               /note="assembly_fragment"
misc_feature   149669..152678
               /note="assembly_fragment"
misc_feature   152779..154:35
               /note="assembly_fragment"
misc_feature   154236..155662
               /note="assembly_fragment"
misc_feature   15576:..1568:9
               /note="assembly_fragment"

BASE COUNT     37089 a   39987 c   40186 g   37757 t   :800 others
ORIGIN

Alignment Scores:
Pred. No.:              1.08e-187        Length:          156819
Score:                  2617.50          Matches:         622
Percent Similarity:     33.13%           Conservative:    1
Best Local Similarity:  39.07%           Mismatches:      2
Query Match:            81.34%           Indels:          969
DB:                     2                Gaps:            4

US-10-029-347-2 (1-625) x AP006283 (1-156819)

Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGlyLeuProAla 20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12647 ATGCTGGCCCAGCCGCAGCGGCTCCTCTTCATCCTGGACGGGGCCGGACGGGCTGCCGGCG 12706

Qy     21 LeuGlyGlyProGluAlaA:aProCysThrAspProPheGluAlaAlaSerGlyAlaAlaArg 40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12707 CTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCGGCGCGG 12766

Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuLeuProThrArgA:aLeuLeuLeuValThrThr 60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13847 CCTGCTGGACAGGCACTGCGGCTGATCAGCTGCATCAGCTGCGATTGGTTGCTGCGCAGGAAGAAG 13906
```

```
Db  12767 GTGCTAGGCGGCTGCTGAGCAAGGCGGCTGCTGCCACGGCCCTCGCTGCTGGTGACCACG 12826
Qy     61 ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGluVal 80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12827 CGGCGCCGCGCCCCCGGAGGCTGCAGGGCCGCCGCCTGTGTTCCCGCAGTGCGCCGAGGTG 12886
Qy     81 ArgGlyPheSerAspLysAspLysLysTyrLysPhePheArgAspGlyArg :00
          |||||||||||||||||||||||||||||||||||||||||||||||||
Db  12887 CGCGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTATTTCCGGGATGAGAGG 12946
Qy    101 ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  12947 AGGGGCCGAGCGGCGGCCTACCGCTTCGTGAAGGAGAACGAGACGCGCTTCGCGCTGTGCTTC 13006
Qy    121 ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuArgGlyLeuGlyArg 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13007 GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGCGGCTCGGTCGG 13066
Qy    141 AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal 160
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13067 GACCTGTCGCGCACGTCCAAGACCACCAGTCAGTGTACCTGCTCTTTCATCACCAGCGTT 13126
Qy    161 LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuAspLeuArgAsnLeuCys 180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13127 CTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACCTGGCGCAATCTGTGC 13186
Qy    18: ArgLeuAlaArgGluGlyValGluValArgArgAlaGlnPheAlaGluLysGluGlu 200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  :3187 CGCCTGGCCCGCGAGGGCGTCCTCGGAGACGCAGGGCCAGTTGCCGAGGAGAACTGGAG :3246
Qy    20: GlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLysGluPro 220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  :3247 CAACTGGAGTTCGTGGCTCCTCCAAAGTGCAGACGCGTGTTCTTTCAGCAAAAAGGAGTGCCG 13356
Qy    221 GlyValLeuGlyThrGlnValThrTyrGlnPheIleAspSerGlnSerPheGlnGlyPheLeu 240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13307 GGCGGTGCTGGGACAGAGGTCACCTACCAGTTCATCGACAGCCAGAGCTTCCAGGAGTTCCTC 13366
Qy    241 AlaAlaLeuSerTyrLeuLeuGluLeuGluSerTyrLeuAspArgThrAlaAlaGlyValGlyVal 260
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13367 GCGGCACTGTCCTACCTGTGGAGAGGACGGGTGCCCAGGAGGACCGGCGTGCGGCCGTT 13426
Qy    261 GlyThrLeuArgGlyAlaAspAlaGlnProHisSerHisLeuValLeuThrThrArgPhe 280
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  :3427 GGGACACTCCTGCTGCTGGGGACGGCCAGCCGACCAGCCACTTGGTGCTCACCACGCGCTTC 13486
Qy    281 LeuPheGlyLeuLeuSerArgAlaLeuArgMetArgAspIleGluArgHisPheGlyCysMet 300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13487 CTCTTCGGACTGCTGAGCCGCGAGGCGCGGATGCGGGACATCGAGCGGCACTTGGCTGCATG 13546
Qy    301 ValSerGluArgValValLysGlnGluGlnGluGluGluGluGlyGlnGlnGlyCys 320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13547 GTTTCAGAGCGCGTGAAGCAGGAGGAGGAGGAGGAGGACAGGGCTGCAGGGACAGGGCTGC 13606
Qy    321 ProGlyValAlaProGluValThrGluGlyValAlaLysLeuAspThrG:uPro 340
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13607 CCCGGAGTGGCACCAGAGGTGACCGAGGGGGCCAAAGGCTTGGACACCGAAGAGCCA 13666
Qy    341 GluGluGluGluGlyGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyr 360
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13667 GAGGAGGAGGAGGGGGAGGAGCCCAACTACCCCACTGGAGTTGCTGTACTGCCTGTAC 13726
Qy    361 GluThrGlnGlnGluAspAlaPheValArgGlnAlaLeuArgPheProGluLeuAlaLeu 380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13727 GAGACCCAGGAGGAGGACGCGTTTGTGCGCCAAGCCCTGCGGTTCCCGGAGCTGGCGCTG 13786
Qy    381 GInArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysCys 400
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13787 CAGCGAGTGCGCTTCTGCCGCATGGACGTGGCGTGTCGTCGGAGGTGCTGC 13846
Qy    401 ProAlaGlyGlnAlaLeuAlaLeuArgLeuIleSerCysArgLeuLysSerArgLeuValAlaAlaGlnGluLysLys 420
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  13847 CCTGCTGGACAGGCACTGCGGCTGATCAGCTGCATCAGCTGCGATTGGTTGCTGCGCAGGAAGAAG 13906
```

Qy          421 LysLysSerLeuGlnLysArgLeuGlnAlaSerLeuGlyGlyGly----------------- 435
13907 AAGAAGAGCGAGCCTGGGGAAGCCGGCTCCAGGCCAGCGAGGTGGCGGCGAGGTGCGTCTCCAGG 13966

Qy          435 ----------------------------------------------------------------- 435
13967 GCAGAGATACTCTCTTGTGTGTGAGGTGTGTGGCCGGTGCAAACCTGTGCACCTCAGTG 14026

Qy          435 ----------------------------------------------------------------- 435
14027 TCCATCTGGCATGGGCGCACTCCAGACACCAGACCCTGGCTCCCCAGATGGGGAGGCTTT 14086

Qy          435 ----------------------------------------------------------------- 435
14087 GACCACCTCCTTGCAAGGCAACCACTGAATTCCTATCTTCTTCCCCAGGCTGTCCTCACC 14146

Qy          435 ----------------------------------------------------------------- 435
14147 ATGCCCACCGTCAGCCACCCACAACCTATGAAAACTTCCTAGTATGGCCCTACAGCTCTG 14206

Qy          435 ----------------------------------------------------------------- 435
14207 GCCCAGCCCTTCTTCCCATTTTACTTTCTGACAATAAGCCTGGAACACACGATGCCCCC 14266

Qy          435 ----------------------------------------------------------------- 435
14267 TGTCCAGAGCCTCTGCACTTTCCCCTTCCCTCCACCTAGAATATTCTGTTTGCTGCTGTT 14326

Qy          435 ----------------------------------------------------------------- 435
14327 CTGCAAGTCTGCCCTGCCTCCCCAGCTGCTGGGGACCCAACTGGCTCCAGAAGTACCC 14386

Qy          435 ----------------------------------------------------------------- 435
14387 TTTCGACCACCCTGCTGTGAATCTCTGCCCCACACCTCCACCAGACCCCTCGGCTCCTCCAG 14446

Qy          435 ----------------------------------------------------------------- 435
14447 GGCAAGGCTTTGCCAGAGTTCCTTTGAATATAGCTCCAATTCAGCCCCTGCCCAGGGGC 14506

Qy          435 ----------------------------------------------------------------- 435
14507 TTGGCATCTGTTGAGAGGATGAATGCCACCGTGTTGGCAGGGGTGTGGCTGGGCTGGGGACCCA 14566

Qy          435 ----------------------------------------------------------------- 435
14567 AAGACCCATGTAGAGGATGTGGGTGATACCGTCCCACTGACCTCACTCAGACACCCCAGT 14626

Qy          435 ----------------------------------------------------------------- 435
14627 GGGATGATGCCTGGGCCACAGATGAGTGGTTTGGTTGGGGGAGGTGTGA 14686

Qy          435 ----------------------------------------------------------------- 435
14687 AGGGGGTGGGCTCAGCAAGGAGGGTGAGGAGGGCAGCTCTGAGACCCCAGGACTACACAG 14746

Qy          435 ----------------------------------------------------------------- 435
14747 ATAGACAGGAGGAGTCCTTTCCGCGCAGCACTGTGAGAGCACAGGTGGGCTTCACCTTCCTCT 14806

Qy          436 ------SerSerGlnGlyThrThrLysGlnLeuProAlaSerLeuLeuLeuHisProLe 452
14807 CTCTCCCAGCAGTTCTCAAGGACCACCAAAACAACTGCCAGCCTCCCTTCTTCATCCACT 14866

Qy          452 uPheGlnAlaMetThrAspProLeuCysHisSerSerLeuThrLeu---------------- 468
            ||||||||||||||||||||||||||||||||||||||||||||
14867 CTTTCAGGCAATGACTGACCCACTGTGCCATCTGAGCCAGCTCACTG-AGTGGCCACAC 14925

Qy          468 ----------------------------------------------------------------- 468
14926 CCCCAGCTCTTCCCACGGAGCGGGCTGAGAGCACAGGTGCCCTGGGCAGAGACGGAGTAT 14985

Qy          468 ----------------------------------------------------------------- 468
14986 GACCTAGGAAAGAAGAGGCTCAGACCAGAGCCTGGGCCTGGTGGGTGCTGGAGAT 15045

Qy          468 ----------------------------------------------------------------- 468
15046 GGCAGGCTCTGCCCACAGAGCAGAGGGAGGGGATGAGGGTCTGGGGCTTCATTAC 15105

Qy          468 ----------------------------------------------------------------- 468
15106 CAGGTTAGGTACTGGTAGGAAAAGGTTGGTCCTGGTGAGTGAAAAGCAGCTAAGCACACT 15165

Qy          468 ----------------------------------------------------------------- 468
15166 GAGGGAGGGATAAAGAGACAAGAGTCTATTTCCTTCTGAAGGGAGGGCTGCCGTTGGA 15225

Qy          468 ----------------------------------------------------------------- 468
15226 AACTGCTGGGTTCTCAGAGCAGTGGAGGAAGGTGTTCAAGCGCCAGGCTCCAAGGTCAGG 15285

Qy          468 ----------------------------------------------------------------- 468
15286 CTGTCTGCAAATGCTAGCCGGCTTCTGCGTGCCAGTAGCAAGTCCATCTGTCGCGAGCTT 15345

Qy          468 ----------------------------------------------------------------- 468
15346 CCATTCCTCCCCTGTGAAATGAGCCAGTCCAGAGTCCAGAGAGTTGGGTGAGATTTAATG 15405

Qy          468 ----------------------------------------------------------------- 468
15406 TGGCTACACATGTACAGTTCTTTTTTTTTTTTTTTTTGAGACGGAGCCTCGCTCTGT 15465

Qy          468 ----------------------------------------------------------------- 468
15466 CGCCCAGGCTGGAGTGCAGTGGTGTGATCTCAGCTCACTGCAGCTCCGCCTCCCGGGTT 15525

Qy          468 ----------------------------------------------------------------- 468
15526 CACGCCCATTCTCCTGCCTCAGCCTCAGCTAGCTGGGACTACAGGTGCCCGCCACTACG 15585

Qy          468 ----------------------------------------------------------------- 468
15586 CCCGCTAATTGTTTGTATTTTAGTAGAGACAGGGTTTCACCGGGTTAGCCAGGATAG 15645

Qy          468 ----------------------------------------------------------------- 468
15646 TCTCGATCTCCTGACCTCGTGATCCACCCGCCTCGCCTCCCAAAGTGCTGGGATCACAG 15705

Qy          468 ----------------------------------------------------------------- 468
15706 GCATGAGCCACCGCGCCCGGCCACACTGTACAGTTCTCACATGCAATAACTTGCCAGTCA 15765

Qy          468 ----------------------------------------------------------------- 468
15766 CTCTTAGCTACCCAAGTGAGCAGAAGAGCAGAAATCTAGGAACTAGTACTGATGG 15825

Qy          468 ----------------------------------------------------------------- 468
15826 TGAAACAGTCCATCCTACGTGATGGCAACTCGGCAAATTAGCTTGATATGGGGGTGGG 15885

Qy          468 ----------------------------------------------------------------- 468
15886 GGATTTGGCAATCTGGACGATAAACATAAGAACCAAAAAGGTGATCACATTTGCCCAGCA 15945

Qy          468 ----------------------------------------------------------------- 468
15946 ATTCTGCTGGTAAAGTAAGATGTTCAGCACACCATTATTTATAATGAAAAAAAAAAG 16005

Qy          468 ----------------------------------------------------------------- 468
16006 GAGAAACAACAACTACCCAACCATTAGGGAATAGCCGGACTAAATAAACTGAGTTGGCC 16065

```
/product="Pyraf5 protein"
/protein_id="AAH31139.1"
/db_xref="GI:21410402"
/db_xref="LocusID:101613"
/translation="MLKSAGTNGPRVQGELRTLCRLAREGILDHHKAQFSEEDLEKLK
LRGSQVOTIFLNKKEIPIGVLKTEVTYQFIDQSFQEFLAALSYLLEAERTGTPAGGVQ
KLLNSDAELRGHLALTTRFLFGLLNTEGLRDIGNHFGCVVPDHVKKDTLRWVOQQSH?
KGPPVGAKKTAELEDIEDAEEEEEEEDLNFGLELLYCLLYETQEEDFVRQALSS2FEI
VLERVRLTRMDLEVLNYCVQQCCPDGQALRLVSCGLVAAKEKKKKSLVKRLKGSQST
KKQ2PVSLLRPLCETWTTPKCHLSVLILSHCRLPDAVCRDLSEA:KVAPLRELGLLQ
SRLTNTGLRLLCEGLAWPKCCQVKTLRMQLPDLQEVINYIVIVLQQSPVITTLDLSGCQ
LPGVIVEPLCAAL.KHPKCSLKTLSLTSVELSENSLRDLQAVKTSKPDLSIIYSK"
```

```
BASE COUNT   884 a   909 c   902 g   759 t
ORIGIN
```

```
Alignment Scores:
Pred. No.:            2.98e-144        Length:        3454
Score:               2020.00          Matches:       4:2
Percent Similarity:  80.03%           Conservative:  65
Best Local Similarity: 69.13%         Mismatches:    :0:
Query Match:         62.77%           Indels:        18
DB:                  10               Gaps:          8
```

US-1C-029-347-2 (1-625) x BC03:13) (:-3454)

```
Qy    31  AspProPheGluAlaAlaSerGlyAlaArgValLeuGlyLeuGlySerLysAlaLeu   50
                                                                    
Db    31  GACCCTTTGGAGGTACACAGTGGCTTGAGAGTGCTGAGCCGGAGGASCTG          90

Qy    51  LeuProThrAlaLeuLeuValThrThrArgAlaAlaAlaAlaProGlyArgLeuGlnGly   70
                                                                      
Db    6:  TTACCAGGAGCTCGCTTGCTAGTGACTACACACGCCATGCCTCTACAGGGAGGTGCAGGC   :50

Qy    71  ArgLeuCysSerProGlnCysA:aGluValArgGlyP:eSerAspLysAspLysLys      90
                                                                       
Cb    15:  AGATTGTGCTGCCCACAGTGCGGCAGAAATACGCGGCTTCCAGACAAGACAAAAAAAG      2:0

Qy    91  TyrPheTyr:LysPhePheA:-gAspGluArgAlaGluArgAlaTyrArgAlaTyrPheVa:Lys   110
                                                                           
Db    211 :ACTTCTCAAGTTCTTCCGGGACGAGCGAGGAGGCAGAGCGCGCTTACCGCCTCGTGAAA   27:

Qy    ::1  GluAsnGluThrLeuPheAlaLeuCysPheValProPheVa:CysTrpIleValCysThr   :30
                                                                        
Db    271 GAGAACGAGACGGCTGTTTGCACTGCCTGCTTGTGCCCTTCGTCTGTGCTGGATCGTGCACA   330

Qy    131 ValLeuArgGlnGlnLeuGluLeuGlyLeuGlyAlaArgAsnLeuCysArgGluLeuArg   :50
                                                                        
Db    33:  GTGCTGCGACAGCAGCTAGAGCTGGGCGGGAATCTCTCCGTACCTCCCAAGACCACTACT   390

Qy    15:  SerValTyrLeuPheIleThrSerVa:LeuSerSeri-aProValAlaAspGlyPrc     170
                                                                       
Db    391 :CTGGTACCTGGTCTTCATCACCAGCATGCTGAAGTCTGCAGGC---ACCAATGGACCC    447

Qy    171 ArgLeuGluGlyAspLeuAr-gAsn:LeuCysArgGlyLeuAlaArgGlu:ValLleuG-y-   189
                                                                        
Db    448 CGGG:TCAGGAGGACGGCTGTGCGAGAAACCTGTGTCGCGGGAGGGGCATCCTGGATCAT   507

Qy    190 ArgArgAlaGlnGlnPheAlaGluLeuGluLysSerGluLeuGluGlnLeuLeuArgGlySerLysVa:   209
                                                                              
Db    508 CATAAAGCACAATTCTCAACAAGAGGACCTGGAGAAATTGAAGCTTCGGGGTTCCCAAGTT   567

Qy    210 GlnThrLeuPheLeuSerLysLysGluLysPheProGlyValLeuProLysLeuGluLeuTyrThrTyr   229
                                                                              
Db    568 CAGACAAATATT:CTCAACAAGAAAGAGATACCAGGAGGTGCTAAAAACTGAGGTCACCTAC   627

Qy    230 GlnPheIleAspSerProPheGlnLeuPheLeuAlaAlaLeu:LeuSerTyrLeuLeuGluAla   249
                                                                           
Db    628 CAGTTCATTGACCAGGAGCTTCCAGGAGTTCTTGCTGCATTGTCATACCTACTACAGAGCT   687

Qy    250 GlyGlyValProArgThrAlaAlaGlyValGlyValThrLeuLeuArgGlyValGlyAspAlaGl:n   269
                                                                             
Db    688 GAGCGAACTCCGGGAAGCCCGCCAGCGCGTGCAAGAGCTCCCTGAACTCGACGCCGGCGAG   747
```

```
Qy   270  ProHisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArg   289
                                                                       
Db   748  CTACGCGGTCATCTTGCACTGACCACCCGGATTCCTCTTTGGCCTGCTAAATACAGAGGGG   807

Qy   290  MetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluAla   339
                                                                       
Db   808  CTTCGTGACATTGAAACCATTTTGGCTGTCGTGTCTGCCAGATCATGTGAAAAGGACACC   867

Qy   310  LeuArgTrpValGlnGlyGlnGly-----GlnGlyCysProGlyValAlaProGluVa:   327
                                                                     
Db   868  CTGCGGTGGGTACAAGGACAGAGACCCACCCCAGTAGGGGCAAAAAAGGAC:          927

Qy   328  ThrGluGlyAlaLysGlyLeuGluAspThrGluProGluGluArgGluArgGluArgGlu   347
                                                                       
Db   928  GCTGAG------------------CTGGAGGACATCAGGACGCCAGAGGAGGAGGAAGAG   975

Qy   348  GluPro---AsnTyrProLeuGluLeuLeuTyrCysLeuTyrGluThrGlnGluAspAla   366
                                                                       
Db   976  GAGGACCTCAATTTGGACTGGAGCTGCTATCCCTGTACCTGGAGGAGGAT          1035

Qy   367  PheValArgGlnAlaLeu:SerProGluLeuAlaLeuGlnArgValArgArgPheCys    386
                                                                     
Db  :036  TTTGTTCGACAGGCTCTCCAGCAGCCTTCCAGAGATAGTACTGGAGCGAGTCAGGTTGACC   1095

Qy   387  ArgMetAspVa:AlaValAlaValAspThrCysValArgCysCysProAlaGlyValGlnAlaLeu   406
                                                                           
Db  1096  CGCATGGACCTTGAAGT:CTGAATTACTGCCTGCGCTGCTGCCCAGATGGCCAGGCTCTG   :155

Qy   407  ArgLeuIleSerCysArgLeuValAlaAlaGlnGlu--------LysLysLysSerLeu   424
                                                                     
Db  1156  AGACTGGTCGAGCTGTGGCG:TGGTGGCGCAAAGGAAGAAGAAGAAGAAGAAGAGGCCTG   1215

Qy   425  GlyLysArgLeuGlnAlaSerLeuGlyGlyGlySerGlnGlyTyrThrArgLysGlnLeu   444
                                                                       
Db  1216  --------GGCTCTCAAAGCACCAAGAAAACACCC              1257

Qy   445  ProAlaSerLeuLeuHisProLeuPheGlAlaMetThrAspProLeuCysHisLeuSer   464
                                                                      
Db  :258  CCAGT:CTCCTTGCTGCGTCCACTCTGTGAGACAATGACCACCCCGAAATGTCATCTGAGT   1317

Qy   465  SerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAspLeuSerGluAla   484
                                                                       
Db  :319  GTTCTGATCTGTCACACTGCAGACTCCCGATGCAGTTTGCCGAGACCTTTCCGAGGCC     :377

Qy   485  LeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuLeuH:sAsnArgLeuSerGluAla   504
                                                                           
Db  1378  CTGAAGGTAGCTCCTGCCTAAGGAGTTGGGCCTCTCCAGAGCCGGCTCACCACACA        :437

Qy   505  GlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgGlnValGlnThrValArg   524
                                                                           
Db  1438  GGCCTGCG:TTACTGTGCAGAAACTGCCAAGTGCCAGGTGAAGACACTCAGG           1497

Qy   525  Va:GlnLeuProAspProGlnArgGlyLeuValGl:yMetLeuArgGlnSer          544
                                                                     
Db  1498  ATGCAGCTGCCTGACCTGCGACCTGATCAATTACTTGGTCATCGTGCTCCAGCAGAGC    1557

Qy   545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr   564
                                                                       
Db  1558  CCAGTCCTGACCACTCTGGACCTCAGCTGGCTGTCAGTCAGTGGCTCATTGTGGAACCT    1617

Qy   565  LeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSerLeuAlaSerVal   584
                                                                       
Db  1618  CTGTGTGCAGCCCTGAAGCACCCTAAATGTCCCAAAGACCCTCAGTCTGACCTCTGTG    1677

Qy   585  GluLeuSerGluGlnSerLeuGlnLeuGlnLeuAlaValLysArgAspLeu           604
                                                                     
Db  1678  GAGCTGAGTGAGAACTCACTGAGGGATCTTCAAGCTGTGAAGACACATCAAAGCCAGATCTG   :737

Qy   605  ValIleThrHisProAlaLeuAspGlyHisProGlnProLysGlu  620
                                                           
Db  1738  TCCATCATATAT---TCAAAATAAGGCACACCTCACCCGGCCTCGGAAG   1782
```

RESULT :4

BC013519
LOCUS        BC013519                2354 bp    mRNA    linear   ROD 20-SEP-2002
DEFINITION   Mus musculus, clone IMAGE:4236272, mRNA, partial cds.
ACCESSION    BC013519
VERSION      BC013519.1  GI:15488764
KEYWORDS     .
SOURCE       Mus musculus (house mouse)
  ORGANISM   Mus musculus
             Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
             Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.
REFERENCE    1  (bases 1 to 2354)
  AUTHORS    Strausberg,R.
  TITLE      Direct Submission
  JOURNAL    Submitted (04-SEP-2001) National Institutes of Health, Mammalian
             Gene Collection (MGC), Cancer Genomics Office, National Cancer
             Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590,
             USA
REMARK       NIH-MGC Project URL: http://mgc.nci.nih.gov
COMMENT      Contact: MGC help desk
             Email: cgapbs-r@mail.nih.gov
             Tissue Procurement: Jeffrey E. Green, M.D.
             cDNA Library Preparation: Life Technologies, Inc.
             cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)
             DNA Sequencing by: Sequencing Group at the Stanford Human Genome
             Center, Stanford University School of Medicine, Stanford, CA 94305
             Web site:      http://www.shgc.stanford.edu
             Contact:    (Dickson, Mark) mcd@paxil.stanford.edu
             Dickson, M., Schmutz, J., Grimwood, J., Rodriguez, A., and Myers,
             R. M.

             Clone distribution: MGC clone distribution information can be found
             through the I.M.A.G.E. Consortium/LLNL at: http://image.llnl.gov
             Series: IRAK Plate: 25 Row: a Column: 10
             This clone was selected for full length sequencing because it
             passed the following selection criteria: Similarity but not
             identity to protein.
FEATURES             Location/Qualifiers
     source          1..2354
                     /organism="Mus musculus"
                     /mol_type="mRNA"
                     /strain="FVB/N"
                     /db_xref="taxon:10090"
                     /clone="IMAGE:4236272"
                     /tissue_type="Kidney, normal. 5 month old male mouse."
                     /clone_lib="NCI_CGAP_Kid14"
                     /lab_host="DH10B"
                     /note="vector: pCMV-SPORT6"
     CDS             <1..1673
                     /codon_start=3
                     /product="Unknown (protein for IMAGE:4236272)"
                     /protein_id="AAH13519.1"
                     /db_xref="GI:15488765"
                     /translation="LLPGARLLVTTRHAATGRLQGRLCSPCCAEIRGFSDKDKKYFF
                     KFFRDERKAERAYRFVKENETLFALCFVPFVCWIVCTVLQQLELGDLSRTSKTTTS
                     VYLLFITSMLKSAGTNGPRVQGELRTLCRLAREGILDHHKAQFSEEDLEKLKLRGSQV
                     QTIFLNKKEIPGV-KTEVTYQFIDQSFQEFLAALSYLLEAERTPGTPAGGVQKLLNSD
                     AELRGHLALTTRFJFGLLNTEGLRDIGNHFGCVVPDJHVKKDTLRWVQGQSHPKGPVG
                     AKKTAELEDIEDAAEEEEEEDINFGELELLYCLYETQEDFVRQALSSLPEIVJERVR
                     LTRMDLEVLNYCVQCCEPDGQALRLVSCGLVAAKEKKKKKSLVKRLKGSQSTKKQPPV
                     SLLRPJCETMTTPKCHLSVLLSHCRLPDAVCRDJSEALKVAPALRELGLLGSRJTNT
                     GLRLLCEGLAWPKCQVKTLRMQLPDLQEVINYLVIVLQQSPVLTTDLSGCCLPGVIV
                     EPICAALKHPKCS.KTLSLTSVELSENSLRDLCAVKTSKPDLSIIYSK"
BASE COUNT      588 a    614 c    542 g    510 t
ORIGIN

Alignment Scores:
Pred. No.:                              Length:        2354
Score:          9.37e-140               Matches:       399
Percent Similarity:   80.24%            Conservative:  64
Best Local Similarity:  69.15%          Mismatches:    96
Query Match:            60.85%          Indels:        18
                                        Gaps:          8

US-10-029-347-2 (1-625) x BC013519 (1-2354)

Qy    50  LeuLeuProThrAlaLeuLeuLeuValThrThrArgAlaAlaAlaAlaProGlyArgLeuGln   69
              |||||||                 ||||||| ||| || ||||||||||||||||||||
Db     3  CTGTTACCAGGAGCTCGCTTGCTAGTGACTACACGCCATGCGCTACAGGGAGGCTGCAG   62

Qy    70  GlyArgLeuCysSerProGlnCysAlaGlyValAlaArgGlyPheSerAspLysAspLysLys   89
              |||||||||||| |||||||||||||||||  |||||||||||||||||||||||||||
Db    63  GGCAGATTGTCTCGCCACCAGTGCGCAGAAATACGCGGCTTCTCAGACAACAAAGACAAAAA   122

Qy    90  LysTyrPhePheTyrLysPheArgGluArgArgAlaGluArgArgAlaTyrArgPheVal   109
              ||||||||||::|||||||||:||||:|||||||||||||||||::|||||||||||
Db   123  AAGTATTTCTTCAAGTTCTCCGGGAGACGAGAGGAAGGCAGAGCGCGCTTACCGCTTCGTG   182

Qy   110  LysGluAsnGluThrLeuPheAlaLeuCysPheValProPheValCysTrpIleValCys   129
              |||||||||||||||||||||||||||||||||||||||||||||::|||||||||
Db   183  AAAGAGAACGAGACGCTGTTTGCACTGTGCTTTGTGCCCTTCGTGTGCTGGATCGTGTGC   242

Qy   130  ThrValLeuArgGlnGlnLeuGluLeuGlyAspLeuSerArgThrSerLysThrThr   149
              |||||||||::|||||||||||||||::||||||:||:|||::||||||::||||
Db   243  ACAGTGCTGCAGCAGCAGTAGCTGGGCCGGATCTCCTGTACCTCCAAGACCACCACT   302

Qy   150  ThrSerValTyrLeuLeuPheIleThrSerMetLeuLysSerAlaGlyThrAsnGlyPro   169
              |||:|||||||||||| ||||||||||||:||||||||||||||||| ::|
Db   303  ACTTCTGTGTACCTGCTCTTCATCACCAGCATGCTGAAGTCTGCAGGC---ACCAATGGA   359

Qy   170  ProArgValGlnGlyGluLeuArgThrLeuCysArgLeuAlaArgGluGlyIleLeuAspHisHis   189
              |||||||||:||||||||||:||||||||||||||||||||||||||||||||||||||
Db   360  CCCCGGGTTCAGGGAGAGCTGAGAACGCTGTGTCGCCTGGCCAGGGAGGGCATCCTGGAT   419

Qy   190  --ArgArgAlaGlnPheAlaGluAlaGluSerLysGluAlaGlnLeuGluLeuGlySerLys   208
                  :::|||||||||:||||||||::|||||:|||||||||:|||||||||||:
Db   420  CATCATAAAGCACAATTCTCAGAAGAGGACCTGGAGAAATTGAAGCTTCGGGGTTCCCAA   479

Qy   209  ValGlnThrLeuPheLeuSerLysSerLysGluLeuProGlyValLeuGluThrGluValThr   228
              :::|||||||||| |||||||||:::||||||||||||||||||||||
Db   480  GTTCAGACACATATTTCTCAACAAGAAAGAGATACCAGGAGTGCTAAAACTGAGGTCACC   539

Qy   229  TyrGlnPheIleAspGlnSerPheGlnGluPheLeuAlaAlaLeuSerTyrLeuLeuGlu   248
              |||||||||| ::|||||||||:|||||||||||||||:::||||||||:
Db   540  TACCAGTTCATTGACCAGAGCTTCCAGGAGTTCTTGGCTGCATTGTCATACCTACTAGAA   599

Qy   249  AspGlyGlyValProArgThrAlaAlaGlyValGlyThrLeuGluArgGlyAspAlaAla   268
              || |||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   600  GCTGGAGCGAACTCCGGGACGCCCCAGCCGGGTGTGCAGAAGTTCCTGAACTCTGACGCG   659

Qy   269  GlnProHisSerHisLeuValLeuThrThrArgPheLeuProHisGlyLeuLeuSerAlaGlu   288
              ::|                       ||||||||||::||||||:||||||||
Db   660  GAGCTACGTGGGTCA TCTTGCACTGACCACCCGGATTCCTCTTTGCCTGTAAATACAGAG   719

Qy   289  ArgMetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGlu   308
              |||||||||||||| || |||||::|||::||||||||||||||||::|||||||||
Db   720  GGGCTTCGTGACATTGGAAACCATTTTGGCTGTGGTGCCAGATCATGTGAAAAAGGAC   779

Qy   309  AlaLeuArgTrpValGlnGlyGlnGly-----GlnGlyCysProGlyValAlaProGlu   326
              :::||| ::|||||||||||||||||            :::|||||||||||||||||
Db   780  ACCCTGCGTGGGTACAAGGACAGAGCCCAGCCGGGCCAAAAAG   839

Qy   327  ValThrGluGlyLeuAlaLysGlyLeuGlnAspThrGluProGluGluGluGluGluGlyGly   346
              ::::::::::::::::::::          :::||||||||||||||||||||||
Db   840  ACTGCTGAG------------------CTGGAGGACATCGAGGACGCGCAGGAGGAGGAGGAA   887

Qy   347  GluGluPro---AsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluAsp   365
              |||:::             ||||||:|||||||||||||||::|||||||||
Db   888  GAGGAGGACCTCAATTTGGACTGGAGCTGCTGTACTGCCTGTATGAGACACAGGAGGAG   947

Qy   366  AlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArgValArgPhe   385
              |||||||||:::          ::||| ||||:::|||||||||||||||:::
Db   948  GATTTTGTTGGACAGGCTCTCAGCAGCCTTCCAGAGATAGTACTGGAGCGAGTCAGGTTG   1007

Qy   386  CysArgMetAspValAlaValLeuSerTyrCysValArgCysCysProAlaGlyValGlnAla   405
              |||||||:::::                         :::::|||||||||||
Db  1008  ACCCGCATGGACCTTGAGGTTCTGAATTACTGCGTGCAGTGCTGCCCAGATGCCAGGCT   1067

```
Qy   406  LeuArgLeuSerCysArgLeuValAlaAlaGlnGlu-----LysLysLysLysSer  423
          ||||||||||:::|||              ||||||||||
Db  1068  CTGAGACTGGTCAGCTGCTGGGCTGGTGGCGGCAAAGGAGGAAGAGGAGAGAGAGC  1127

Qy   424  LeuGlyGlyLysArgLeuGlnAlaSerLeuGlyGlySerGlySerGlyThrGlyThrLysGlyGln  443
          ||||||                              ||:||
Db  1128  CTGGTGAGCGGCTGAAG-------------GGCTCTCAAAGCACCAAGAAACAA  :169

Qy   444  LeuProLeuAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLeu  463
          ||||||||||   ||| :::  |||||
Db  :170  CCCCAGTCTCCTTGCTGCCACTCTGTGAGACAATGACTACCCGAAATGTCATCTG  1229

Qy   464  SerSerLeuThrLeuSerHisCysLysLysLeuProAspAlaValCysArgAspLeuSerGlu  483
          ||||  |:||||||||||||||:|||||
Db  1230  AGTGTTCTGATCTGTCACACTGCAGACTCCCTGATGCAGTTGCCGAGACCTTTCCGAG  1289

Qy   484  AlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGlu  503
          ||||||||||||||||||||:||||||||||||
Db  :290  GCCCTGAAGGTAGCTCCTGCCCTTAAGGGAGTTGGGCCTTCCTCCAGAGCCGGCTCACCAAC  1349

Qy   504  AlaGlyLeuArgMetLeuSerGluGluValAlaTrpProGlnCysArgValGlnThrVal  523
          ||||||    :||        :::||||||      ||||||||
Db  :290  ACAGGCCTGCGTTTACTGTGCGAAGGCCTGGCTTGGCCCCAAGTGCCAGGTGAAGACACTC  1409

Qy   524  ArgValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGln  543
          ||:::|||    ||| :::   ||||||||||
Db  :410  AGGATGCAGCTGCCTGACCCCAAGAGGGTGATCAATTACTTGGTCATCGTGCTCCAGCAG  1469

Qy   544  SerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThr  563
          ||||||||||||||     :::  ||:||||||
Db  1470  AGCCCAGTCCTGACCACCTGGACCCTCAGTGCCTGTCAGCTGCCTGGCGTCATTGTGGAA  :529

Qy   564  TyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuAspLeuAlaSer  583
          ||||||||||||||||||||:|||||||||
Db  1530  CCCTGTGTGGCAGCCGTGCTAAAGCACCCTAAATGCTCCCTAAAGACCCTCAGTCTGACCTCT  1589

Qy   584  ValGluLeuSerGluGlnSerLeuGlnGluGlnGlnLeuGlnGluLysAspAlaValLysArgAlaLysAlaLysProAspAsp  603
          |||||||||||||||||||    :::  ||:||||:||||
Db  1590  GTGGAGCTGAGTGAGAACTCACTGAGGGATCTTCAAGCTGTGAAGACATCAAAGCCAGAT  1649

Qy   604  LeuValIleThrHisProAlaLeuAspGlyHisProGlnHisProProLysGlu  623
          ||||| ::: |||
Db  1650  CTGTCCATCATATAT----TCAAAATAAGGCACACCTCACCCGGCCTCGGAG  1697
```

RESULT 15
AC132968
LOCUS      AC132968          238533 bp    DNA    linear    HTG 20-NOV-2002
DEFINITION Rattus norvegicus clone CH230-172C3, *** SEQUENCING IN PROGRESS
           ***, 4 unordered pieces.
ACCESSION  AC132968
VERSION    AC132968.4  GI:25139219
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ENRICHED.
SOURCE     Rattus norvegicus (Norway rat)
ORGANISM   Rattus norvegicus
           Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
           Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae;
           Rattus.
REFERENCE  1  (bases 1 to 238533)
AUTHORS    Muzny,D.Marie., Metzker,M.Lee., Abramzon,S., Adams,C., Aider,J.,
           Allen,C., Allen,H., Alsbrooks,S., Amin,A., Anguiano,D.,
           Aryaiebechi,V., Aoyagi,A., Ayodeji,M., Baca,E., Baden,H.,
           Baldwin,D., Bandaranaike,D., Barber,M., Barnstead,M., Benahmed,F.,
           Biswalo,K., Blair,J., Blankenburg,K., Blyth,P., Brown,M.,
           Bryant,N., Buhay,C., Burch,P., Burrell,K., Calderon,E.,
           Cardenas,V., Carter,K., Cavazos,I., Ceasar,H., Center,A.,
           Chacko,J., Chavez,D., Chen,G., Chen,R., Chen,Y., Chen,Z., Chu,J.,
           Cleveland,C., Cockrell,R., Cox,C., Coyle,M., Cree,A., D'Souza,L.,
           Davila,M.L., Davis,C., Davy-Carroll,L., De Anda,C., Dederich,D.,
           Delgado,O., Denson,S., Deramo,C., Ding,Y., Dinh,H., Divya,K.,
           Draper,H., Dugan-Rocha,S., Dunn,A., Durbin,K., Duval,B., Eaves,K.,
           Egan,A., Escotto,M., Eugene,C., Evans,C.A., Falls,T., Fan,G.,
           Fernandez,S., Finley,M., Flagg,N., Forbes,L., Foster,M., Foster,P.,

           Fraser,C.M., Gabisi,A., Ganta,R., Garcia,A., Garner,T., Garza,M.,
           Gebregeorgis,E., Geer,K., Gill,R., Grady,M., Guerra,W., Guevara,W.,
           Gunaratne,P., Haaland,W., Hamil,C., Hamilton,C., Hamilton,K.,
           Harvey,Y., Havlak,P., Hawes,A., Henderson,N., Hernandez,J.,
           Hernandez,R., Hines,S., Hladun,S.L., Hodgson,A., Hoques,M.,
           Hollins,B., Howells,S., Hulyk,S., Hume,J., Idlebird,D., Jackson,A.,
           Jackson,L., Jacob,L., Jiang,H., Johnson,B., Johnson,R., Jolivet,A.,
           Karpathy,S., Kelly,S., Kelly,S., Khan,Z., King,L., Kovar,C.,
           Kowis,C., Kraft,C.L., Lebow,H., Levan,J., Lewis,L., Li,Z., Liu,J.,
           Liu,J., Liu,W., Liu,Y., London,P., Longacre,S., Lopez,J.,
           Lorensuhewa,L., Louiseged,H., Lozado,R.J., Lu,X., Ma,J.,
           Maheshwari,M., Mahindartne,M., Mahmoud,M., Malloy,K., Mangum,A.,
           Mangum,B., Mapua,P., Martin,K., Martin,R., Martinez,E.,
           Mawhiney,S., McLeod,M.P., McNeill,T.Z., Meenen,E.,
           Milosavljevic,A., Miner,G., Minja,E., Montemayor,J., Moore,S.,
           Morgan,M., Morris,K., Morris,S., Munidasa,M., Murphy,M., Nair,L.,
           Nankervis,C., Neal,D., Newton,N., Nguyen,N., Norris,S.,
           Nwokelemeh,O., Okwuonu,G., Olarpunnagoon,A., Pai,S., Parks,K.,
           Pasternak,S., Patl,H., Perez,A., Perez,L., Pfannkoch,C.,
           Plopper,F., Poindexter,A., Popovic,D., Primus,E., Pu,L.-L.,
           Puazo,M., Quiroz,J., Rachlin,E., Reeves,K., Regier,M.A., Reigh,R.,
           Reilly,B., Reilly,M., Ren,Y., Reuter,M., Richards,S., Riggs,F.,
           Rives,C., Rodkey,T., Rojas,A., Rose,M., Rose,R., Ruiz,S.J.,
           Sanders,W., Savery,G., Scherer,S., Scott,G., Shatsman,S., Shen,H.,
           Shetty,J., Shvartsbeyn,A., Sisson,I., Sitter,C.D., Smajs,D.,
           Sneed,A., Sodergren,E., Song,X.-Z., Sorelle,R., Sosa,J.,
           Steimle,M., Strong,R., Sutton,A., Svatek,A., Tabor,P., Taylor,C.,
           Taylor,T., Thomas,N., Thomas,S., Tingey,A., Trejos,Z., Usmani,K.,
           Valas,R., Vera,V., Villasana,D., Waldron,D., Walker,B., Wang,J.,
           Wang,Q., Wang,S., Warren,J., Warren,R., Wei,X., White,F.,
           Williams,G., Willson,R., Wleczyk,R., Wooden,H., Worley,K.,
           Wright,D., Wright,R., Wu,J., Yakub,S., Yen,J., Yoon,L., Yoon,V.,
           Yu,F., Zhang,J., Zhou,J., Zhou,X., Zhao,S., Dunn,D., von
           Niederhauserr,A., Weiss,R., Smith,D.R., Holt,R.A., Smith,H.O.
           Weinstock,G., and Gibbs,R.A.
TITLE      Direct Submission
JOURNAL    Unpublished
REFERENCE  2  (bases 1 to 238533)
AUTHORS    Rat Genome Sequencing Consortium.
TITLE      Direct Submission
JOURNAL    Submitted (05-SEP-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
REFERENCE  3  (bases 1 to 238533)
AUTHORS    Rat Genome Sequencing Consortium.
TITLE      Direct Submission
JOURNAL    Submitted (20-NOV-2002) Human Genome Sequencing Center, Department
           of Molecular and Human Genetics, Baylor College of Medicine, One
           Baylor Plaza, Houston, TX 77030, USA
COMMENT    On Nov 20, 2002 this sequence version replaced gi:23264838.
           The sequence in this assembly is a combination of BAC based reads
           and whole genome shotgun sequencing reads assembled using Atlas
           (http://www.hgsc.bcm.tmc.edu/projects/rat/). Each contig described
           in the feature table below represents a scaffold in the Atlas
           assembly (a 'contig-scaffold'). Within each contig-scaffold,
           individual sequence contigs are ordered and oriented, and separated
           by sized gaps filled with Ns to the estimated size. The sequence
           may extend beyond the ends of the clone and there may be sequence
           contigs within a contig-scaffold that consist entirely of whole
           genome shotgun sequence reads. Both end sequences and whole genome
           shotgun sequence only contigs will be indicated in the feature
           table.
           --------------- Genome Center
              Center: Baylor College of Medicine
              Center code: BCM
              Web site: http://www.hgsc.bcm.tmc.edu/
              Contact: hgsc-help@bcm.tmc.edu
           --------------- Project Information
              Center project name: KBMS
              Center clone name: CH230-172C3
           --------------- Summary Statistics
              Assembly program: Phrap; version 0.990329
              Consensus quality: 223317 bases at least Q40

```
         Consensus quality: 225463 bases at least Q30
         Consensus quality: 226655 bases at least Q20
         Estimated insert size: 230654; sum-of-contigs estimation
         Quality coverage: 5x in Q20 bases; sum-of-contigs estimation
*    NOTE: Estimated insert size may differ from sequence length.
         (see http://www.hgsc.bcm.tmc.edu/docs/Genbank_draft_data.html).
*    NOTE: This is a 'working draft' sequence. It currently
         consists of 4 contigs. The true order of the pieces
         is not known and their order in this sequence record is
         arbitrary. Gaps between the contigs are represented as
         runs of N, but the exact sizes of the gaps are unknown.
*    This record will be updated with the finished sequence
*    as soon as it is available and the accession number will
*    be preserved.
*            1    222461: contig of 222461 bp in length
         222462    222561: gap of unknown length
         222562    224949: contig of 2388 bp in length
         224950    225049: gap of unknown length
         225050    228242: contig of 3193 bp in length
         228243    228342: gap of unknown length
         228343    238533: contig of 10191 bp in length.

FEATURES             Location/Qualifiers
     source          1..238533
                     /organism="Rattus norvegicus"
                     /mol_type="genomic DNA"
                     /db_xref="taxon:10116"
                     /clone="CH230-172C3"
     misc_feature    1..2349
                     /note="wgs contig"
     misc_feature    124155..209379
                     /note="clone_boundary
                     clone_end:T7
                     site"
                     end sequence:BH290527"
BASE COUNT      61451 a    53521 c    52385 g    60005 t    11471 others
ORIGIN

Alignment Scores:
Pred. No.:              238533
Score:                  1746.50
Percent Similarity:     36.80%
Best Local Similarity:  32.68%
Query Match:            54.27%

Length:         238533
Matches:        452
Conservative:   57
Mismatches:     113
Indels:         763
Gaps:           11
```

```
Db  172318 TTGCAAGGCAGCCGTCTTATTCCTTCTATTCCTTCTTCAAATGACCTCACCACGAGTACTCTCAGTCAT 172377
Db  :72318 TTGCAAGGCAGCCGTCTCTATTCCTTCTTCAAATGACCTCACCACGAGTACTCTCAGTCAT 172377

Cy     463 -------------------------------------------------              463
Cy     423 -----------------------SerLeuGlyLysArgLeuGlnAlaSer-              43:
                                    |||   |:|
Db  173516 TTTCTATTACTGTGAGCTATCCTCAGTGAGCGAAGAGTAAGGACAATTACAGGAACCAA 173516
Db  172378 GCAGATCAAGGCAACCCTCCAAGTTCAGCAGGTAGTCCACGAAG-CAACCGAGCCCC :72436

Qy     463 -------------------------------------------------              463
Qy     431 -------------------------------------------------              431
Db  173576 TAGAGGCAGGCCATGTCGAAGCAGCAGCAGTCAGGGATGATGATGAGGATGAGGAG :73576
Db  172437 CTCTCTGATGCACTTTCTCATACATTCTGGAATATGCCATGTCCTCTGTGTGCTGAGCGT 172496

Cy     463 -------------------------------------------------              463
Cy     431 -------------------------------------------------              43:
Db  173636 GAAAACTGGCTCTGTGGGGGTCAGGGCCCAGCACGTGAAAATAAAACAACGGAGCCGA 173636
Db  :72497 TGTACCTGGCGCGTCCCTCCATCCCAACACTATTTGCTGCTGTCTTCTTCAATGCTACTCTG 172556

Qy     463 -------------------------------------------------              463
Qy     431 -------------------------------------------------              431
Db  :73637 AAAGGCAGTCGCATGGGCTTCAGGAATTCTCACCAGAAAACCAAGTATACAGAGAGGTGCTCA :73696
Db  172557 CCTCCACCACTTGGGTCCAACTACTGGCCCCAAGGCACCCCCATTGACCACATTATTCCC 172616

Qy     463 -------------------------------------------------              463
Qy     43:  -------------------------------------------------              431
Db  173696 CTTCATATTTGTAATGGGAAAGAATAAATAAATCAGAAATAAATAGCAGAGCTGTTCAGCCAC 173756
Db  :72617 ATATCTTCTTCCTTGCCCA:CGTCCACTGA::CGTCCAGGTTCACCGACATGGGCATTTGTT :72676

Qy     463 -------------------------------------------------              463
Qy     43:  -------------------------------------------------              431
Db  173756 TGAGGATGACTGCTCAGCCACTGGGATGGATGCTGCTCAGCCACTGGGTATGACTGCTCAGC 173816
Db  172677 TTTTATGAGGTCAGCTCTACACAATGCTTGTCTGTCCCAAGGCTCGCCCACTGTTGAGAGTATA 172736

Qy     463 -------------------------------------------------              463
Qy     43:  -------------------------------------------------              43:
Db  :73876 CACTGGGGATGACTGCTCTCAGCCACTAGGGATGACTGACTGCTCAGCCACAGGGATGACTGCTC :73876
Db  172737 CACCAGGATAATGGCAGGACCTGGTCCTGGGATCTCGCAGCCCACACTTGGGTTTTGCTC 172796

Qy     463 -------------------------------------------------              463
Qy     432 -----LeuGlyGly-----------------------------------              434
                  ::|:|:|:|
Db  173935 AGCCACTGGGGGATGACTGCTGCTCAGCCATTGGGATGCAGCCCATCCGAACCAAATT 173935
Db  172797 AAA:TGGTTSGGGGAGGAGGAGGAGGGATGCAAGGGATG:GAGATAAATGA:TGTTC:GGAACCTCAGG 172856

Qy     463 -------------------------------------------------              463
Qy     434 -------------------------------------------------              434
Db  :73937 ATGAAGAACTGTAGATTTCTGACTGAACA:CAGTCTGCCAGTGAATAGCTAGT:TGTTC 173996
Db  :72857 AAGGATTGAGGCCTTCCAGAAAGGCCTGGAAGGGGTGCTACAGACGTCAAACCTTGTTC: 172915

Cy     463 -------------------LeuSe 469
Cy     464 -------------------SerSerLeuThr--------------------              45:
                          :::|||||||:
Db  173997 AGCTGGAGGTAGAGAAGTGGCCAGTGTTACAGTCTTCCATGTCTTCTTGACCACAGCCTGTC 174056
Db  172917 TTC:CCCCTCCAGTTCTCAAAGCACACCGGGAAACAACCCCCAGCCTCCTTCCTGCTGCGCCA :72576

Qy     469 rHisCysLysLeuProAspAlaValCysArgAspLeuSerGluALeuArgALAalaPr 489
Qy     452 LeuPheGlnAlaMet--------------------------------------              452
             :::|||||||||||:::|||                |||  ||:|||||
Db  174116 ACACTGCAAACTCCCTGATGCAGTTTGTCGAGACCTTTCTGAGGCTCTGAAGGTAGCTCC 174116
Db  172977 CTCTGTGAGGCAATGATTACCCAGCAATGTGGTCTGGTCTGAGTATCTGACGTGAGTCACCATG 173036

Cy     489 oAlaLeuThrGluGlyLeuLeuHisAsnArgLeuse=GluAlaGlyLyLeuArgMetLe 509
Cy     456 --------------------ThrAspProLeuCysHisLeu--              463
             |::||||||||||                     |||
Db  174176 TTCCCTAAGGGAGCTGGGCCTCCCTCCAGAACCGACTCACTGAGGCAGGCCTGCGTTTACT 174176
Db  173037 CTCTGGAGGGGATTTACACCTGGGTGGGTCTGGACAGGGACAGAATTTGCCTCTTAGGGAA :73096

Qy     509 uSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrVal------------ 523
Qy     456 -------------------------------------------------              456
Db  174236 AAGCCAAGGCCTGGCCTTGGGCCAAATGCAGAGGTGCAGACACTCAGGTAAGGCCTGGTTT: 174236
Db  173097 GGGTAAGGACATAGGCCTACCTGGCTTTCACCACTAT:TGGGGCTGTCAGGTGAGGGCTG :73156

Qy     524 --Ar 524
Qy     456 ::|:              456
Db  :74237 GGAAGGATGGGGTAGGGAGGAACCATCCCAACAGCAACCCCGAGTTGGCTTTCCCCACAG 174296
Db  :73157 GTGCAAGACCT:GTGGGAATCAGAAGCTGCACTGCCTTTGGGACACGGCGAGAAGCA 173216

Qy     524 gValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSe 544
Qy     456 -------------------------------------------------              456
             :::|||::::||             ||:         :::|||:::|||
Db  174356 GATACAGATGCCTGGCCTGCCTCCAAGAGGTGATCCACCTGGTCATTGTGCTCCAGCAGAG 174356
Db  173217 ACAGTGTTTTCCTTTTGAAGACAGCTCCTAAGTCTGAAGGACACTGCCCGGCCACCCCT :73276

Qy     544 rProAlaLeuThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTy 564
Qy     457 -----------------ThrAspProLeuCysHisLeu--              463
             |||||::::||                     |||
Db  174416 CCCAGTCCTAACCACTCTGGACCTCCAGTGGCTGTCCAGCTGCTGTGGAGGGTGGGGAACC 174416
Db  173277 ATGCAGGAAAGTGTTGAGGGAGCTCAGAAGCTGGTCAGAAGCTGGTGTGTGCCTTGCA 173336

Qy     564 rLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeu------------ 579
Qy     463 -------------------------------------------------              463
             |||||::::||
Db  174475 TCTGTGTTCAGCCCTGAAGCACCCTAAATGGGCCTAAAGACCCT-CAGTAAGGCAGG 174475
Db  173337 GCAAGTCAGTCTGCCCATCCCACTCGTGTCGTCCCCTGTAACATACCCAGCTGA:CCCAGCT :73396

Qy     579 -------------------------------------------------              579
Qy     463 -------------------------------------------------              463
Db  174535 GGGACCCATGCTGCTTGCTGAACCTGAGTGGGAGAAGTGCTTTGGGCG 174535
Db  173397 CACAGGATTAGTGTAAGATTTACTGAGGCTCCACGGAGA:AGTTCTTAAGATATGCTAGG 173456
```

```
Qy    579 ------------------------------------------------------------ 579

Db 174536 GCGTCCTCTCTGAGACCTCACACAGAGATCACACAGATCATGATTGAGCTCACACAGCTCTG 174595

Qy    579 ------------------------------------------------------------ 579

Db 174596 GGTGACCAGCCCATAAAGAGCCTACTTCTGCAGGCCCATCCAAAGCTCCTGCTCCTCAAA 174655

Qy    579 ------------------------------------------------------------ 579

Db 174656 CAAAGGGCCCTAAGCTGTGGGCACAGTCCGAGCATAGCCCTCTCTGTATCCCCCCAGAG 174715

Qy    579 ------------------------------------------------------------ 579

Db 174716 GGCCTCAGCTTTTAATAGACCCCAACCGACTGCCTCCTTGTACTTGTTATATCCTAAGCA 174775

Qy    579 ------------------------------------------------------------ 579

Db 174776 AGCAGTGCCAACCACCCCCGCTCCACCTCAAGCCCTGGGTTCCCCCACAGGCAGTCGGC 174835

Qy    580 -----------------------------------------SerLeuAlaSe 583
                                                     |||||
Db 174836 CTTTGCCCCTGACTGATCCTTGTTCCCCCTGTCTTACACATGTGGCTGCAGTCTGACTTC 174895

Qy    583 rValGluLeuSerGluGlnLeuGlnLeuGlnArgAlaValLysArgAlaLysProAs 603
              ||||||||||| ||||:::||||| ||| |||: |||   ||
Db 174896 TGTGGAGCTGACTGAGAATCCACTGAGAGAGCTTCAAGCTGTGAAGACATTAAAGCCAGA 174955

Qy    603 pLeuValIleThrHisProAlaLeuAspGlyHisProGlnProLysGluLeuIleSe 623
              ||| ||||||            ||||||  ||||||
Db 174956 TCTGGCCATCATCACATTCAAAATTGGGCACACATCCTCAGCCCTCAGAAGGGATGAAGCAG 175015

Qy    623 rThrPhe 625
              |||
Db 175016 TGTCTTC 175022
```

Search completed: October 3, 2003, 15:49:12
Job time : 6649 secs

OM protein - nucleic search, using frame_plus_p2n model

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

Run on: October 3, 2003, 09:15:40 ; Search time 442 Seconds
(without alignments)
38:7.079 Million cell updates/sec

Title: US-10-029-347-2
Perfect score: 3218
Sequence: 1 MLAQPQRLLFLLDGADLLPA.........ITHPALDGHPCPPKELIST 625

Scoring table: BLOSUM62
Xgapop 10.0 , Xgapext 0.5
Ygapop 10.0 , Ygapext 0.5
Fgapop 6.0 , Fgapext 7.0
Delop 6.0 , Delext 7.0

Searched: 252756 seqs, 134971901 residues

Total number of hits satisfying chosen parameters: 5105512

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Command line parameters:
-MODE=frame_p2n.model -DEV=xih
-Q=/cgn2.1/USPTO_spool/US10029347/runat 02102003 173436 17066/app.query.fasta_1.775
-DB=N Geneseq_19Jun03 -QFMT=fastap -SUFIX=rng -MINMATCH=0.1 -LOOPCU=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cdi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=ext -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=200000000
-USER=US10029347 @CGN_1_1_312 @runat 02102003 173436 17066 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database : N Geneseq 19Jun03:*
1: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1980.DAT:*
2: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1981.DAT:*
3: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1982.DAT:*
4: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1983.DAT:*
5: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1984.DAT:*
6: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1985.DAT:*
7: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1986.DAT:*
8: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1987.DAT:*
9: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1988.DAT:*
10: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1989.DAT:*
11: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1990.DAT:*
12: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1991.DAT:*
13: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1992.DAT:*
14: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1993.DAT:*
15: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1994.DAT:*
16: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1995.DAT:*
17: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1996.DAT:*
18: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1997.DAT:*
19: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1998.DAT:*
20: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA1999.DAT:*
21: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2000.DAT:*
22: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001A.DAT:*
23: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2001B.DAT:*
24: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2002.DAT:*
25: /SIDS1/gcgdata/geneseq/geneseqn-embl/NA2003.DAT:*

and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | ID | Description |
|---|---|---|---|---|---|---|
| 1 | 3218 | 100.0 | 2689 | 24 | ABS63485 | DNA encoding human |
| 2 | 3218 | 100.0 | 3365 | 22 | AAD08836 | Human G-protein co |
| 3 | 3170 | 98.5 | 2763 | 24 | ABS63486 | Human leucine-rich |
| 4 | 3144 | 97.7 | 2109 | 24 | ABX97042 | Human NOV18a cDNA. |
| 5 | 2882 | 89.6 | 5556 | 24 | AAL47132 | Pyrin domain conta |
| 6 | 2882 | 89.6 | 6461 | 22 | AAS00033 | Human AT-AS-2-enco |
| 7 | 2239 | 69.6 | 1795 | 22 | ABA08979 | Human vasopressin |
| 8 | 2209 | 68.6 | 2054 | 24 | ABS63487 | Human leucine-rich |
| 9 | 1737 | 54.0 | 2251 | 14 | AAQ38723 | AII/AVPv2 receptor |
| 10 | 1243 | 38.6 | 933 | 22 | AAK83113 | Human immune/haema |
| 11 | 1167.5 | 36.3 | 989 | 22 | AAK62741 | Human immune/haema |
| 12 | 1147.5 | 35.7 | 990 | 24 | ABL90582 | Human polynucleoti |
| 13 | 826.5 | 25.7 | 3300 | 24 | AAL47129 | Pyrin domain conta |
| 14 | 825.5 | 25.7 | 3186 | 22 | AAL44363 | Human NB-ARC and C |
| 15 | 800.5 | 24.9 | 3857 | 22 | AAD14323 | Human PYRIN 1 (PYR |
| 16 | 800.5 | 24.9 | 3857 | 25 | ABX93556 | Huma cDNA encoding |
| 17 | 791 | 24.6 | 704 | 24 | ABQ29972 | Oligonucleotide fo |
| 18 | 791 | 24.6 | 704 | 24 | ABQ29973 | Oligonucleotide fo |
| 19 | 789 | 24.5 | 704 | 24 | ABQ29970 | Oligonucleotide fo |
| 20 | 789 | 24.5 | 704 | 24 | ABQ29971 | Oligonucleotide fo |
| 21 | 762 | 23.7 | 4422 | 22 | AAD02760 | Human NB-ARC and C |
| 22 | 762 | 23.7 | 4556 | 22 | AAD02764 | Human NAC beta iso |
| 23 | 762 | 23.7 | 5122 | 22 | AAD12951 | Human G-protein co |
| 24 | 762 | 23.7 | 5523 | 22 | AAF83651 | Human CARD-7 polyp |
| 25 | 762 | 23.7 | 5523 | 25 | ABS55497 | cDNA encoding huma |
| 26 | 762 | 23.7 | 5523 | 25 | ABS56030 | cDNA encoding huma |
| 27 | 760.5 | 23.6 | 2847 | 22 | AAS01487 | Human secreted pro |
| 28 | 760.5 | 23.6 | 2847 | 25 | ABZ73494 | Secreted protein-e |
| 29 | 758 | 23.6 | 5100 | 22 | AAL47127 | Pyrin domain conta |
| 30 | 716 | 22.2 | 4466 | 22 | AAD02765 | Human NAC gamma or |
| 31 | 712.5 | 22.1 | 4194 | 22 | AAD02761 | Human NB-ARC and C |
| 32 | 712.5 | 22.1 | 4329 | 22 | AAD02762 | Human NB-ARC and C |
| 33 | 694 | 21.6 | 4931 | 24 | ABL59333 | Nucleotide sequenc |
| 34 | 691 | 21.5 | 2763 | 22 | ABQ75801 | Human MDDT-13 enco |
| 35 | 680.5 | 21.1 | 678 | 25 | ACA56413 | Norway rat signal |
| 36 | 660 | 20.5 | 3085 | 23 | AAS68111 | DNA encoding novel |
| 37 | 650 | 20.2 | 3263 | 24 | AAI70684 | Human nucleotide b |
| 38 | 650 | 20.2 | 3459 | 24 | ABQ78049 | Human CGDD encodin |
| 39 | 635 | 19.7 | 3069 | 24 | AAL47134 | Pyrin domain conta |
| 40 | 634 | 19.7 | 3150 | 22 | AAI67184 | Nucleotide sequenc |
| 41 | 619.5 | 19.3 | 3431 | 22 | AAD14322 | Human nucleotide b |
| 42 | 619.5 | 19.3 | 3431 | 25 | ABX93555 | Huma cDNA encoding |
| 43 | 616.5 | 19.2 | 3189 | 24 | AAI67185 | Nucleotide sequenc |
| 44 | 616.5 | 19.2 | 3531 | 24 | AAL47128 | Pyrin domain conta |
| 45 | 603.5 | 18.8 | 2218 | 20 | AAZ20641 | CBDAX01 coding se |

ALIGNMENTS

RESULT 1
ABS63485
ID AB563485 standard; cDNA; 2689 BP.
XX
AC ABS63485;
XX
DT 15-NOV-2002 (first entry)
XX
DE DNA encoding human leucine-rich repeat small intestine I (HLRRSI1).
XX
KW Human; human leucine-rich repeat small intestine I; HLRRSI1; asthma;
KW proliferative disorder; gastrointestinal disorder; renal disorder;
KW neural disorder; reproductive disorder; calcium regulation; apoptosis;
KW immune system; anaemia; human immune deficiency virus; HIV; cancer;
KW blood coagulation disorder; autoimmune disorder; allergic reaction;
KW inflammatory condition; cardiovascular disorder; ischaemia;
KW neurological disorder; infectious disease; cytokine production;

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,

KW  expressed sequence tag; EST; gene; ss.
XX
OS  Homo sapiens.
XX
PN  WO200261086-A2.
XX
PD  08-AUG-2002.
XX
PF  20-DEC-2001; 2001WO-US49739.
XX
PR  22-DEC-2000; 2000US-257774P.
XX
PA  (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI  Feder J.  Ramanathan C.  Mintier G;
XX
DR  WPI; 2002-619252/66.
DR  P-PSDB; ABG78454.
XX
PT  New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT  their fragments and homologues, useful for preventing, treating and
PT  ameliorating medical conditions, e.g. proliferative, gastrointestinal,
PT  or renal disorders -
XX
PS  Claim 1; Figure 1; 336pp; English.
XX
CC  The invention relates to isolated nucleic acid molecules (I) encoding
CC  human leucine-rich repeat small intestine I (HLRRSI1) polypeptides.
CC  The nucleic acid molecules and polypeptides are useful for preventing,
CC  treating and ameliorating medical conditions, such as proliferative,
CC  gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC  related to aberrant calcium regulation or apoptosis modulation, either
CC  directly or indirectly. They are also useful for treating, preventing
CC  and/or diagnosing diseases, disorders and/or conditions of: immune system
CC  by activating or inhibiting the proliferation, differentiation, or
CC  mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC  ataemia; immunologic deficiency syndromes, e.g. human immure deficiency
CC  virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC  e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC  myasthenia gravis; asthma or allergic reactions; inflammatory
CC  conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC  e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC  ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC  Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC  pneumonia, or viral, bacterial, and fungal infections. The HLRRSI1
CC  polypeptides are useful for modulating cytokine production, antigen
CC  presentation, or other processes such as boosting immune responses.
CC  ABS63485-ABS63504 represent HLRRSI1 coding sequences and PCR primers of
CC  the invention.
XX
SQ  Sequence 2689 BF; 534 A; 861 C; 844 G; 450 T; 0 other;

Alignment Scores:
Pred. No.:              1.1e-269
Score:                  3218.00
Percent Similarity:     100.00%
Best Local Similarity:  100.00%
Query Match:            24

Length:        2689
Matches:       625
Conservative:  0
Mismatches:    0
Indels:        0
Gaps:          0

US-10-029-347-2 (1-625) x ABS63485 (1-2689)

Qy      1 MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db     75 ATGCTGGCCCAGCCGCAGCGGCTCTTCATCCTGGACGGCGCGGACGAGCTGCCGGCG  134

Qy     21 LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    135 CTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCG  194

Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuSerLeuProThrAlaLeuLeuLeuValThrThr  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    195 GTGCTAGGCGGGGCTGCTGAGTAAGGGCGCGCTGCTGCCACGGCCCTCCTTGCTGTGACCACG  254

ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuArgCysSerProGlnCysAlaGlyVal  80
Qy   61
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  255 CGGGCCGCCGCCCCGGGAGGCCTGCAGGGCCGCCTGCGTTCCCCGCAGTGCGCCGGAGTG  3:4

ArgGlyPheSerAspLysAspLysLysTyrPheArgPheArgAspGl:Arg  100
Qy   81
         ||||||||||||||||||||||||||||||||||||||||||||||||
Db  315 CGCGGCTTCTCCGACAAGGACAAGAAGTATTTCAGAGTTCTCCGGGA:GAGAGG  374

ArgAlaGluAlaArgAlaTyrArgPheValLysGl:uAsnGluThrLeuPheAlaLeuCysPhe  120
Qy  101
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  375 AGGGCCGAGGCACGCGCCTACCGCTTCGTGAAGGAGGAAACGACCCTGT:CGGCTGTGCTTC  434

ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGluGlyGlyArg  140
Qy  121
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  435 GTGCCCTTCGTGTGTGTGATCGTGTGCACCGTGCTGCGCCAGCAGTGGAGCTCGGTCGG  494

AspLeuSerArgThrSerLysThrThrSerValTyrLeuPheIleThrSerVal  160
Qy  141
         |||||||||||||||||||||||||||||||||||||||||||||||||||
Db  495 GACCTGTCGCGCACGTCCAAGACCACCAGCGTCTACCTGTTTCATCACCAGCGTT  554

LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuAspArgAsnLeuCys  180
Qy  161
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  555 CTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGC  614

ArgLeuLeuAlaArgGluGlyValArgValArgArgAlaGlnPheAlaGluGluGlu  200
Qy  181
         |||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  615 CGCCTGGCCCGCGAGGGCGTCCTCGGCACGCAGGGCCAGTTGCCGGAGAAGGAACTGAG  674

GlnLeuGluLeuArgGlySerGlyValGlnThrArgLeuPheLeuSerLysLysGluLeuPro  220
Qy  201
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  675 CAACTGGAGCTTCGTGGCTCCAAGTGCAGACGCGTTTCTCAGCAAAAGAGCTGCCG  734

GlyValLeuGluThrGlyValThrTyrGlnPheIleAspSerGlnProPheGlnGluLeu  240
Qy  221
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  735 GGCGTGCTGGAGACAGGAGTCACCTACCAGTTCATCGACAGCCAGAGTTCCTC  794

AlaAlaLeuSerTyrLeuGluLeuAspArgAlaAlaGlyGlyValProArgArgThrAlaAlaGlyValVal  260
Qy  241
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  795 GCGGCACTGTCCTACCTGCTGGAGGACGGGCGGGCGCCCAGGACGCGCCTGGCGGCGGTT  854

GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrThrArgPhe  280
Qy  261
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  855 GGGACACTTCCTGCGTGGGACGGCCCAGCCACCACTTGGTGCTCACCACGCGCTTC  914

LeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPheGlyCysMet  300
Qy  281
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  915 CTC::CGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGGCCACTTCGGCTGCATG  974

ValSerGluArgValLysGlnGluValGlyAla:LeuArgTrpValGlnGlyValGlyGlnCys  320
Qy  301
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  975 GTTTCAGAGCGTGTGAAGCAGGAGGTGGGTGCGAGGACCAGGGACAGGGCTGC  1034

ProGlyValAlaProGluValThrGluLeuGlyAlaLysGlyLeuLeuAspThrGluProPro  340
Qy  321
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1035 CCCGAGTGGCACCAGAGTGACGAGGGGCCAAAGGGCTCGGAGGACACCGAAGAGCCA  1094

GluGluGluGluGluProGluProPheGluProAsnTyrProLeuGluLeuTyrCysLeuTyr  360
Qy  341
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1095 GAGGAGGAGGAGGAGCCGGAGCCCTTCGAGCCGAACTACCCACTGGAGTTGCTACTGCCTGTAC  1154

GluThrGlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu  380
Qy  361
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1155 GAGACGCAGGAGGACGCGTTGTGCGCCAAGCCCTGTCGCGTTCCGGAGCTGGCGCTG  1214

GlnArgValArgPheCysProArgMetAspValAlaValAlaLeuSerTyrCysValArgCysCys  400
Qy  381
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1215 CAGCGAGTGCGCTTCTGCCCATGGACGTGGCTGTTCTGAGCTACTGCGTGAGGTGCTGC  1274

ProAlaGlyGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnAlaGluLysLys  420
Qy  401
         |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1275 CCTGCTGGACAGGCGGCCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCCAGGAGAAG  1334

```
Qy  421 LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGlyGlySerSerGlnGlyThr 440
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1335 AAGAAGAGCCTGGGGAAGCGGCTCCAGGCAGGCGGCTCTGGTGGGCGGCAGTTCTCAAGGCACC 1394

Qy  441 ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlrAlaMeThrAspProLeu 460
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1395 ACAAAACAACTGCCAGCCCTCCTTCTTCATCCACTCTTTCAGGCAATGACTGACCCACTG 1454

Qy  461 CysHisLeuSerSerLeuThrLeuSerHisCysLysLysProAspAlaValCysArgAsp 480
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1455 TGCCATCTGAGCAGCCTCACGCTGTCCCACTGCAAACTCCTGACGCGGTCTGCCGAGGAC 1514

Qy  481 LeuSerGluAlaLeuAlaGlyAlaAlaProAlaAlaLeuThrGluGlyLeuLeuHisAsnArg 500
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1515 CTTTCTGAGGCCCTGGCGGGCNGCCCCGCACTGACGGAGCTGGGGCCTCTCCACAACAGG 1574

Qy  501 LeuSerGluAlaGlyLeuLeuArgMetLeuSerGluGluGlyLeuAlaTrpProGlnCysArgVal 520
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1575 CTCAGTGAGGCGGGACTGCTGATGCTGAGTGAGGAGGGCCTAGCCTGGCCCCAGTGCAGGGTG 1634

Qy  521 GlxThrValArgValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMet 540
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1635 CAGACGGTCAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAGTACCTGGTGGGGATG 1694

Qy  541 LeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaPro 560
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1695 CTTCGCCAGAGCCCCGCCCTGACCACCCTGGATCTCAGCGGGTCCCAACTGCCCGCCCCC 1754

Qy  561 MetValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSer 580
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1755 ATGGTGACCTACCTGTGCGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGACCCTCAGT 1814

Qy  581 LeuAlaSerValGluLeuSerGluGlnLeuSerLeuLeuGlnLnAlaValLysArgAla 600
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1815 CTGGCCTCTGTGGAGCTGAGCGAGCAGCTCAGTCACTACAGGAGCTTCAGGGTCTGGAAGAGCA 1874

Qy  601 LysProAspLeuValIleThrHisProAlaLeuAspAspGlyHisProGlnProProLysGlu 620
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1875 AAGCCGGATCTGGTCATCACACACCCAGGCTGGACGCGGACGCCCACAACCTCCCAAGGAA 1934

Qy  621 LeuIleSerThrPhe 625
        |||||||||||||||
Db 1935 CTCATCTCGACCTTC 1949
```

RESULT 2

AAD08836

ID   AAD08836 standard; cDNA; 3365 BP.
XX
AC   AAD08836;
XX
DT   04-SEP-2001 (first entry)
XX
DE   Human G-protein coupled receptor-2 (GCREC-2) cDNA.
XX
KW   Human; G-protein coupled receptor-2; GCREC-2; gene therapy; cirrhosis;
KW   transgenic animal; proliferative disorder; actinic keratosis; hepatitis
KW   nephrotropic; cancer; breast; bladder; bone marrow; brain; uterus;
KW   leukaemia; adenocarcinoma; lymphoma; melanoma; myeloma; epilepsy; stroke;
KW   neurological disorder; Alzheimer's disease; Parkinson's disease; nausea;
KW   Huntington's disease; multiple sclerosis; dementia; angina pectoris;
KW   central nervous system disorder; cardiovascular disorder; hypertension;
KW   atherosclerosis; congestive heart failure; gastrointestinal disorder;
KW   dysphagia; peptic oesophagitis; spasm; gastritis; anorexia; pyrosis;
KW   pancreatitis; Crohn's disease; diarrhoea; autoimmune disorder; anaemia;
KW   inflammatory disorder; Acquired Immune Deficiency Syndrome; AIDS;
KW   Addisor's diseases; allergy; asthma; diabetes mellitus; antithyroid;
KW   atopic dermatitis; glomerulonephritis; Grave's disease; psoriasis;
KW   rheumatoid arthritis; ulcerative colitis; osteoporosis; antifungal;
KW   metabolic disorder; obesity; nootropic; protozoacide; virucide; ss.
XX
OS   Homo sapiens.
XX
FH   Key                 Location/Qualifiers

FT   CDS                 1..2673
FT                       /tag= a
FT                       /product= "Human GCREC-2 protein"
XX
PN   WO200142288-A2.
XX
PD   14-JUN-2001.
XX
PF   07-DEC-2000; 2000WO-US33382.
XX
PR   10-DEC-1999; 99US-0172852.
PR   22-DEC-1999; 99US-0171732.
PR   14-JAN-2000; 2000US-0176148.
PR   21-JAN-2000; 2000US-0177331.
XX
PA   (INCY-) INCYTE GENOMICS INC.
XX
PI   Burford N, Baughn MR, Au-Young J, Yang J, Lu DAM, Reddy R;
XX
DR   WPI; 2001-381635/40.
DR   P-PSDB; AAE04546.
XX
PT   New human G-protein coupled receptor polypeptides for diagnosing,
PT   preventing, and treating cell proliferative, neurological,
PT   cardiovascular, gastrointestinal, autoimmune and metabolic disorders
XX
PS   Claim 5; Page 159-160; 175pp; English.
XX
CC   The present sequence is human G-protein coupled receptor-2 (GCREC-2)
CC   cDNA. GCREC is useful in somatic or germline gene therapy to correct a
CC   genetic deficiency, to express a conditionally lethal gene product and
CC   to express a protein which affords protection against intracellular
CC   parasites and also for diagnosis of disorders associated with expression
CC   of GCREC. GCREC is also useful for generating hybridisation probes useful
CC   in mapping the naturally occurring genomic sequences and to create
CC   knockin humanised animals (pigs) or transgenic animals (mice or rats) to
CC   model human diseases. GCREC is used to diagnose, prevent and treat
CC   proliferative disorders (actinic keratosis, arteriosclerosis, cirrhosis,
CC   hepatitis and cancer); cancer (breast, bladder, bone marrow, brain,
CC   uterus cancer, leukaemia, adenocarcinoma, lymphoma, melanoma and myeloma)
CC   neurological disorders (epilepsy, stroke, Alzheimer's, Huntington's,
CC   Parkinson's disease, multiple sclerosis, dementia and other central
CC   nervous system disorders); cardiovascular disorders (angina pectoris,
CC   hypertension, atherosclerosis, congestive heart failure);
CC   gastrointestinal disorders (dysphagia, peptic oesophagitis, oesophageal
CC   spasm, gastritis, gastric carcinoma, anorexia, nausea, abdominal angina,
CC   pyrosis, pancreatitis, Crohn's disease, diarrhoeal); autoimmune/
CC   inflammatory disorders (acquired immunodeficiency syndrome (AIDS),
CC   Addison's diseases, allergies, anaemia, asthma, diabetes mellitus, atopic
CC   dermatitis, glomerulonephritis, Grave's disease, osteoarthritis,
CC   psoriasis, rheumatoid arthritis, ulcerative colitis, bacterial, fungal,
CC   parasitic, protozoal and helminthic infections) and metabolic disorders
CC   (obesity, osteoporosis, viral infections).
XX
SQ   Sequence 3365 BP; 612 A; 1106 C; 1114 G; 533 T; 0 other;

Alignment Scores:

| | | | |
|---|---|---|---|
| Pred. No.: | 1.5e-269 | Length: | 3365 |
| Score: | 3218.00 | Matches: | 625 |
| Percent Similarity: | 100.00% | Conservative: | 0 |
| Best Local Similarity: | 100.00% | Mismatches: | 0 |
| Query Match: | 100.00% | Indels: | 0 |
| DB: | 22 | Gaps: | 0 |

US-10-029-347-2 (1-625) x AAD08836 (1-3365)

```
Qy   1 MetLeuAlaGlnProGlnArgLeuLeuPheLeuLeuAspGlyAlaAspGluLeuProAla 20
       ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 799 ATGCTGGCCCAGCCCCAGCGCGGCTGCTGTTCATCCTGACGGACGGGGCAGACGAGCTGCCCGCG 858

Qy  21: LeuGlyGlyProGluAlaAlaProCysThrArgProPheGluAlaAlaSerGlyAlaArg 40
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 859 CTGGGGGGGCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCCCGG 918
```

PF    20-DEC-2001; 2001WO-US49739.
XX
PR    22-DEC-2000; 2000US-257774P.
XX
PA    (BRIM ) BRISTOL-MYERS SQUIBB CO.
XX
PI    Feder J, Ramanathan C, Mintier G;
XX
DR    WPI; 2002-619252/66.
XX
PT    New isolated nucleic acid molecules encoding HLRRSI1 polypeptides, or
PT    their fragments and homologues, useful for preventing, treating and
PT    ameliorating medical conditions, e.g. proliferative, gastrointestinal,
PT    or renal disorders -
XX
PS    Example 1; Page 221-222; 336pp; English.
XX
CC    The invention relates to isolated nucleic acid molecules (I) encoding
CC    human leucine-rich repeat small intestine 1 (HLRRSI1) polypeptides.
CC    The nucleic acid molecules and polypeptides are useful for preventing,
CC    treating and ameliorating medical conditions, such as proliferative,
CC    gastrointestinal, renal, neural, or reproductive disorders; or disorders
CC    related to aberrant calcium regulation or apoptosis modulation, either
CC    directly or indirectly. They are also useful for treating, preventing
CC    and/or diagnosing diseases, disorders and/or conditions of: immune system
CC    by activating or inhibiting the proliferation, differentiation, or
CC    mobilisation of immune cells; haematopoietic cells e.g. thrombocytopenia,
CC    anaemia; immunologic deficiency syndromes, e.g. human immune deficiency
CC    virus (HIV) infection, HTLV-BLV infection; blood coagulation disorders,
CC    e.g. arterial thrombosis; autoimmune disorders, e.g. Addison's disease,
CC    myasthenia gravis; asthma or allergic reactions; inflammatory
CC    conditions, e.g. chronic prostatitis, sepsis; proliferative disorders,
CC    e.g. cancer; cardiovascular disorders, e.g. arrhythmia, myocardial
CC    ischaemias, aneurysms; neurological disorders, e.g. Alzheimer's disease,
CC    Huntington's chorea; infectious diseases, e.g. measles, mumps,
CC    pneumonia, or viral, bacterial and fungal infections. The HLRRSI1
CC    polypeptides are useful for modulating cytokine production, antigen
CC    presentation, or other processes such as boosting immune responses.
CC    ABS63485-ABS63504 represent HLRRSI1 coding sequences and PCR primers of
CC    the invention.
XX
SQ    Sequence 2763 BP; 544 A; 890 C; 870 G; 459 T; 0 other;

Alignment Scores:
Pred. No.:              2763
Score:          1.7e-265
Percent Similarity:     3270.00
Best Local Similarity:  96.00%
Query Match:            96.00%
                        98.51%
                        24

Matches:        2763
Conservative:   624
Mismatches:     0
Indels:         1
Gaps:           26
                1

US-10-029-347-2 (1-625) x ABS63486 (1-2763)

Qy     1  MetLeuAlaGlnProGlnArgLeuPheLeuLeuProGlyAlaAspGlyAlaAspGluProAla   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    75  ATGCTGGCCCAGCCCCAGCGCCTGTTCTTCATCCTGGACGGCGCCGACGGAGCTGCCGGGCG  134

Qy    21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg     40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   135  CTGGGGGGCCCCGAGGCGGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCGGCGG  194

Qy    41  ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr   60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   195  GTGCTAGGCGGGCTGCTGAGTAAGGCGCTGCTCCCCACGGCGCCCTCCTGCTGGTGACCACG  254

Qy    61  ArgAlaAlaAlaProGlyArgLeuGlnArgLeuCysSerProGlnCysAlaGluVal     80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   255  CGCGCCGCCGCCCCCGGGAGGCTGCAGCGGCTGTGTTCCCCGCAGTGCGCCGAGGTG  314

Qy    81  ArgGlyPheSerAspLysAspLysLysLysTyrPhePheLysLysPheArgAspGluArg  100
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   315  CGGGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGAGG  374

Qy   101  ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe  120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   375  AGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC  434

Qy   121  ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGlnLeuGlyArg     140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   435  GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGGCAGCTGCAGCTCGGGCGG     494

Qy   141  AspLeuSerArgThrSerThrThrSerValTyrLeuPheIleThrSerVal         160
          |||||||||||||||||||||||||||||||||||||||||||||||||||
Db   495  GACCTGTCGCGCACCAGTACCACCAGTGTACCTGCTTTTCATCACCAGCGTT         554

Qy   161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCys  180
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   555  CTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGC  614

Qy   181  ArgLeuAlaArgGluGlyValValGlyArgArgAlaGlnAlaGluLysGluLeuGlu     200
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   615  CGCCTGGCCCGCGAGGGCGTCCTCCGACGCAGGGCCCAGTTGCCGAGAAGGAACTGGAG     674

Qy   201  GlnLeuGluArgGlySerLysLysValGlnThrLeuPheLeuSerLysLysGluLeuPro  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   675  CAACTGGAGCTTCGTGGCTCCAAGTGCGAGACGCTGTTTCTCAGCAAAAAGGAGCTGCCG  734

Qy   221  GlyValLeuGluThrGluValThrTyrGlnPheIleAspAspAspGlnSerPheGlnPheLeu  240
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   735  GGGGTGCTGGAGACAGAGGTCACCTACCAGTTCATCGACGACCAGAGCTTCCAGGAGTTCCTC  794

Qy   241  AlaAlaLeuSerTyrLeuGluGluGluThrAlaLeuGlyGlyVal           260
          |||||||||||||||||||||||||||||||||||||||||||||
Db   795  GCGGCACTGTCCTACCTGGAGGAGGAGACGGCGGGGGGCGGGCGTT           854

Qy   261  GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisValLeuValLeuThrThrArgPhe  280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   855  GGGACACTCCTGCGTGGGGACGCGCAGCCGCACAGCCACTTGGTGCTCACCACGCGCTTC  914

Qy   281  LeuPheGlyLeuLeuSerArgAlaGlnArgMetAspAspIleGluArgHisPheGlyCysMet  300
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   915  CTCTTCGGACTGCTGAGCCGCGGCGGATGGCGACGCATGGAGCCCCACTTCGGCTGCATG  974

Qy   301  ValSerGlnArgValLysGlnLnUAlaLeuArgTrpValGlnGlnGlyGlnCys     320
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   975  GTTTCAGAGCGCGTGTGAAGCAGCAGGAGGCCCTGCGGTGGGTGCAGCAGGGACAGGCTGC  1034

Qy   321  ProGlyValAlaProGluValThrGluGluThrGlyValLeuGluAspThrGluGluPro  340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1035  CCCGGAGTGGCACCAGAGGTGACCGAGGGGCAAAGGCTCGAGGACACCGAAGAGCCA  1094

Qy   341  GluGluGluGluGluGluGluGluGluGluProGluValArgPheValArgProGluLeuAlaLeu  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1095  GAGGAGGAGGAGGAGGAGGAGCCAGAGGTGCGGTTCGTGCGGCCTGAGTTGGCTCTGTACTGCC  1154

Qy   361  GluThrArgGlnAspAlaProPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu  380
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1155  GAGACGCGAGGCAGGACGCGGTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCGCTG  1214

Qy   381  GlnArgValArgPheCysArgMetAspAspValAlaLeuSerTyrCysValArgCys     400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1215  CAGCGAGTGCGCTTCTGCCGCATGGACGTGGCTGTTTCTGAGCTACTGCGTGAGGTGCTGC  1274

Qy   401  ProAlaGlyGlnAlaLeuArgLeuuSerCysArgLeuuValAlaAlaGlnLnUuLysLys  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1275  CCTGCTGGACAGGCACTGCGGCTGTCGTGCCGGCTGGTTGCTGCTGCGCAGGAGAAGAAG  1334

Qy   421  LysLysSerLeuGlyLysArgLeuGlnAlaSerLeuuGlyGlySerGlnGlyThr  440
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1335  AAGAAGAGCCTGGGGAAGCGGCTCCAGGCCAGCCTGGGTGGCGGCAGTTCTCAAGGCACC  1394

Qy   441  ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeu  460
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1395  ACAAAACACTGCCAGCTCCCCTTCTTCATCCACCTTTCAGGCAATGACTGACCCCACTG  1454

Qy   461  CysHisLeuSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgArgAsp  480

```
Db            1455
                   ||||||||||||||||| |||||||||||| ||||||||||  ||
Qy   481  LeuSerGluAlaLeuArgAlaaAlaProAlaLeuThrG:uLeuGlyLeuLeuHisAsnArg  500
Db   1515  C-TTCTGAGCCCTGAGGGACTGACGGGCCAGCCCCGACTGACGGAGCTGGGCCTCCTCCACAACAGG  1574
                   |||||||||  ||| |||||  |||||  ||||||  |||  ||| |||
Qy   501  LeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpPro:nCysArgVal  520
Db   1575  CTCAGTGAGGCGGGACTGCTATGCTGAGGTGAGGCCTAACCTGGCCGCAGTGCAGGGTG  1634
Qy   521  GlnThrVal-  523
Db   1635  CAGACCGGT-CAGGTGAGGCCTGGCCTGGGAGGACCGTGGCGTGCCCCGCCACCCCAGC  1693
Qy   524  ------ArgValGlnLeuProAspProGlrArgG:yLeuGln  535
Db   1694  AGCTCCTGAGGTGCGGCCCTCCCACAGGGTACAGCTGCCTGACCCCCAGCGAGGGCTCCAG  1753
Qy   536  TyrLeuVa:GlyMetLeuArgGlnSerProAlaLeuThrLeuAspLeuSerGlyCys  555
Db   1754  TACCTGGTGGGTATGCTGCGGCAGAGCCCCGCCCTGACCACCCTGGATCTCAGCGGCTGC  1813
Qy   556  GlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGly  575
Db   1814  CAACTGCCCGCCCCCATGGTGACCTACCTGTGTGCAGTCCTGCAGCACCAGGATGCGGC  1873
Qy   576  LeuGlnThrLeuSerLeuAlaSerValGluLeuSerG:uGlnSerLeuGlnGluLeuGln  595
Db   1874  C-GCAGACCCTCAGTCTGGCCTCTGTGGAGCTGAGCGAGCAGTCACTACAGGAGCTTCAG  1933
Qy   596  AlaValLysArgGlnGlnGlnLysProLeuValIleThrHisProAlaLeuAspGlyHisPro  615
Db   1934  GCTGTGAAGAGAGCAAAGCCCGATCTGGTCATCACACCAGCCCTGGACGGCGGGCCACCCA  1993
Qy   616  GlnProProLysGluLeuIleSerThrPhe  625
                   ||||||||||||||||||||||||||||||
Db   1994  CAACCTCCCAAGGAACTCATCTCGACCTTC  2023
```

RESULT 4
ABX97042
```
ID   ABX97042 standard; cDNA; 2109 BP.
XX
AC   ABX97042;
XX
DT   19-SEP-2002;
DT   20-MAY-2003 (first entry)
XX
DE   Human; NOV18a cDNA.
XX
KW   NOVX; cytostatic; cardiant; antiarteriosclerotic; antiasthmatic; cancer;
KW   hypotensive; cardiomyopathy; bronchial asthma; gene therapy; vaccine;
KW   human; gene; ss.
XX
OS   Homo sapiens.
XX
PN   WO200272757-A2.
XX
PD   19-SEP-2002.
XX
PF   08-MAR-2002; 2002WO-US06908.
XX
PR   08-MAR-2001; 2001US-274101P.        PR   20-MAR-2001; 2001US-277239P.
PR   08-MAR-2001; 2001US-274194P.        PR   20-MAR-2001; 2001US-277321P.
PR   08-MAR-2001; 2001US-274281P.        PR   20-MAR-2001; 2001US-277327P.
PR   08-MAR-2001; 2001US-274322P.        PR   21-MAR-2001; 2001US-277791P.
PR   09-MAR-2001; 2001US-274849P.        PR   22-MAR-2001; 2001US-277833P.
PR   12-MAR-2001; 2001US-275235P.        PR   23-MAR-2001; 2001US-278152P.
PR   13-MAR-2001; 2001US-275578P.        PR   26-MAR-2001; 2001US-278894P.
PR   13-MAR-2001; 2001US-275579P.        PR   27-MAR-2001; 2001US-278999P.
PR   13-MAR-2001; 2001US-275601P.        PR   27-MAR-2001; 2001US-279036P.
PR   14-MAR-2001; 2001US-276000P.        PR   28-MAR-2001; 2001US-279344P.
PR   16-MAR-2001; 2001US-276776P.        PR   30-MAR-2001; 2001US-277338P.
PR   19-MAR-2001; 2001US-276994P.        PR   30-MAR-2001; 2001US-279995P.
                                         PR   30-MAR-2001; 2001US-280233P.
                                         PR   02-APR-2001; 2001US-280802P.
                                         PR   02-APR-2001; 2001US-280822P.
                                         PR   02-APR-2001; 2001US-280900P.
                                         PR   04-APR-2001; 2001US-281194P.
                                         PR   13-APR-2001; 2001US-283675P.
                                         PR   30-APR-2001; 2001US-287424P.
                                         PR   02-MAY-2001; 2001US-288066P.
                                         PR   03-MAY-2001; 2001US-288342P.
                                         PR   03-MAY-2001; 2001US-288528P.
                                         PR   15-MAY-2001; 2001US-291190P.
                                         PR   16-MAY-2001; 2001US-291099P.
                                         PR   16-MAY-2001; 2001US-291240P.
                                         PR   30-MAY-2001; 2001US-294485P.
                                         PR   31-MAY-2001; 2001US-294889P.
                                         PR   31-MAY-2001; 2001US-294899P.
                                         PR   18-JUN-2001; 2001US-299027P.
                                         PR   19-JUN-2001; 2001US-299303P.
                                         PR   19-JUN-2001; 2001US-299310P.
                                         PR   10-JUL-2001; 2001US-304354P.
                                         PR   31-JUL-2001; 2001US-309198P.
                                         PR   16-AUG-2001; 2001US-312903P.
                                         PR   10-SEP-2001; 2001US-318462P.
                                         PR   12-SEP-2001; 2001US-318770P.
                                         PR   27-SEP-2001; 2001US-325430P.
                                         PR   27-SEP-2001; 2001US-325681P.
                                         PR   18-OCT-2001; 2001US-330380P.
                                         PR   31-OCT-2001; 2001US-335301P.
                                         PR   14-NOV-2001; 2001US-332172P.
                                         PR   14-NOV-2001; 2001US-332271P.
                                         PR   14-NOV-2001; 2001US-332272P.
                                         PR   04-NOV-2001; 2001US-333184P.
                                         PR   14-NOV-2001; 2001US-333272P.
                                         PR   21-NOV-2001; 2001US-332094P.
                                         PR   03-DEC-2001; 2001US-337426P.
                                         PR   03-DEC-2001; 2001US-338092P.
                                         PR   04-DEC-2001; 2001US-337185P.
                                         PR   03-JAN-2002; 2002US-345705P.
                                         PR   07-MAR-2002; 2002US-092900.
XX
PA   (CURA-) CURAGEN CORP.
XX
PI   Padigaru M, Spytek KA, Shenoy SG, Taupier RJ, Pena CEA, Li L;
PI   Zerhusen BD, Gusev V, Ji W, Gorman L, Miller CE, Kekuda R;
PI   Patturajan M, Gangolli E, Vernet CAM, Guo X, Tcherhev V;
PI   Fernandes ER, Casman SZ, Malyankar UM, Gerlach V, Liu Y;
PI   Anderson D, Spaderna SK, Catterton E, Burgess C, Leite M, Zhong H;
PI   Alsobrook JP, Lepley DM, Rieger DK;
XX
DR   WPI; 2002-723332/78.
DR   P-PSDB; ABU65075.
XX
CC   NOVX polypeptides and polynucleotides, useful for preventing or
CC   treating a disorder associated with aberrant NOVX expression or
CC   activity e.g., cancer, hypertension, atherosclerosis, cardiomyopathy or
CC   bronchial asthma -
XX
PS   Claim 13; Page 149; 1103pp; English.
XX
CC   This invention describes novel human NOVX polypeptides which have
CC   cytostatic, cardiant, antiarteriosclerotic, antiasthmatic and
```

```
CC   hypotensive activity. Pharmaceutical compositions comprising the NOVX
CC   proteins or nucleic acid molecules or NOVX antibodies are useful for
CC   preventing or treating a disorder associated with aberrant NOVX
CC   expression or activity e.g. cancer, hypertension, atherosclerosis,
CC   cardiomyopathy or bronchial asthma. The products of the invention can
CC   be used for gene therapy or in a vaccine. ABX97008-ABX97185 are cDNA
CC   fragments amplified and isolated by the PCR primers and probes
CC   represented in ABX13460-ABX13462 and ABX97186-ABX97593. ABX97008-ABX97185
CC   encode the NOVX proteins described in ABU65041-ABU65218.
XX
SQ   Sequence 2109 BP; 371 A; 673 C; 717 G; 348 T; 0 other;

Alignment Scores:
Pred. No.:
Score:                 3144.00        Length:          2109
Percent Similarity:    98.40%         Matches:          615
Best Local Similarity: 98.40%         Conservative:       0
Query Match:           97.70%         Mismatches:         4
DB:                                   Indels:             6
                       24             Gaps:               1

US-10-C29-347-2 (1-625) x ABX97042 (1-2109)
```

```
Qy    1 MetLeuAlaGlnProGlnArgLeuGluPheIleLeuAspAspGlyAlaAspGluProAla   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197 ATGCTGGCCCAGCCGCAGCGGCTGGAGTTCATCCTCGACGACGGCGCCGACGAGCTGCCGGCG 256

Qy   21 LeuGlyGlyTyrProGluLysAlaAlaProCysThrAspProPheGluAlaAlaSerGlyValAlaArg   40
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257 CTGGGGGGCTACCCCGAGAAGGCGGCGCCCTGCACAGACCCCTTCGAGGCGGCAAGCGGCGTGGCGCGG 316

Qy   41 ValLeuGlyLysLeuGlyLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr   60
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  317 GTGCTAGGCGGCGGCTGCTGAGCAAGGCGCTCCTGCCCACGGCCCTCCTGCTGGTGACCACG 376

Qy   61 ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysSerAlaGlyVal   80
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  377 CGGGCCGCCGCGCCCCGGGAGGCCGCAGGGCCGGCCGCTGTGTTCCCGCAGTGCGCCGAGGTG 436

Qy   81 ArgGlyTyrPheSerAspLysAspLysLysLysTyrPheThrTyrLysSerTyrPheArgAspAspLysuArg  100
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  437 ACAAAACAACTGCCAGCCTCCTTCTTCATCCACTCTTCACAAGTTCTTCCGGGATGAGAGG 496

Qy  101 ArgAlaGluArgAlaAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe  120
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  497 AGGGCGAGGGCGGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC 556

Qy  121 ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluGluGluGlyArg  140
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  557 GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCTGGAGGAGGAGGGGCGG 616

Qy  141 AspLeuSerArgThrSerThrThrSerValTyrLeuLeuPheIleThrSerVal  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  617 GACCTGTCGCGCACAGTCCAAGACCACCAGCGTCAGTGTACCTGTTTTCATCACCAGCGTT 676

Qy  161 LeuSerSerArgAlaProAlaAlaSerGlyTyrProArgLeuGlnGlyArgLeuArgAsnLeuCys  180
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  677 CTGAGCTCGGCTCGGCTCCGGCCGCCAGCGGGCTACCCGCGGCTGCAGGGCCGGTTGCGCAATCTGTGC 736

Qy  181 ArgLeuAlaArgGlyValLeuGlyValLysArgArgAlaGlnProAlaGluAlaLysGlyLeuGlu  200
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  737 CGGCTGGCGCGCGGCGTCCTCGGACGTGAAGCGCAGGGCGCAGCCTGCCGAGGCGAAGGGACTGGAG 796

Qy  201 GlnLeuGluLeuArgGlyGlySerLysSerValGlnThrLeuLeuPheLeuSerLysLysGluLysGluPro  220
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  797 CAACTGGAGCTTCGTGGCGGCAGACGCGTGTTTCAGACGCTGCTGTTTCTCAGCAAAAAGGAGGTGCCG 856

Qy  221 GlyValLeuGluThrGluValThrThrTyrGlnPheIleAspProPheGlnSerPheGlnHisProLeuLeu  240
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  857 GGCGTGCTGGAGACAGAGGTCACCACCTACCAGTTCATCGACCCGTTCCAGAGCTTCCTTC 916

Qy  241 AlaAlaLeuSerTyrLeuGluGluArgGlyAspGlyGlyValProArgThrAlaAlaGlyValVal  260
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  917 GCGGCACTGTCCTACCTGGAGGAGAGGGGACGGGGCGGGGTGCCCAGGACCGCGGCGGGGGTT 976
```

```
Qy  261 GlyThrLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrThrArgPhe  280
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  977 GGGACACTCCTGCGTGGGGACGCCCAGCCGCACAGCCACTGGTGCTCACCACGCGCTTC 1036

Qy  281 LeuPheGlyLeuLeuSerAlaGluArgMetArgAspIleGluArgHisPheGlyCysMet  300
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1037 CTCTTCGGACTGCTGAGCGCGGAGCGGATGCGCGACATCGAGCGGCACTTCGGCTGCATG 1096

Qy  301 ValSerGluArgValLysGlnGluAlaLeuArgTrpValGlnGlyValGlnGlyCys  320
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1097 GTTTCAGAGCGTGTGAAGCAGGAGGCCCTGCGGTGGGTGCAGGGAGGACAGGGCTGC 1156

Qy  321 ProGlyValAlaProGluValThrGlyValGluLysGlyLeuGluAspThrGluGluPro  340
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1157 CCCGGAGTGGCCACCAGAGGTGACCGGGGTGGAAAGGGCCTGGAGGACACCGAAGAGCCA 1216

Qy  341 GluGluGluGluGluGluGluGluGlyGluAlaArgPheGluValArgGlnAlaLeuTyrCysLeuTyr  360
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1217 GAGGAGGAGGAGGGAGGAGGAGAGGAGCCCAACTACCCCACTGGAGTTGCTGACTGCCTGTAC 1276

Qy  361 GluThrGlnGluGluAspAlaPheAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu  380
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1277 GAGACGCAGGAGGACGCGTTTGTGCGCCAAGCCCTGTGCCGGTTCCCGGAGCTGGCGCTG 1336

Qy  381 GlnArgValArgPheCysArgMetAspAspValAlaValAlaValLeuSerTyrCysValArgCysCys  400
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1337 CAGCGGGTGCGCTTCTGCCGCATGGACGACGTGGCGGTGGCTGTCCTGTCGAGGTGCTGC 1396

Qy  401 ProAlaGlyGlnAlaAlaLeuArgLeuSerCysArgGlyLeuValAlaAlaGlnGluGluLysLys  420
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1397 CCTGCTGCACAGGCACTGCGGCTGATCAGCTGCAGATTGGTGCTGCGCAGGAGGAAGAAG 1456

Qy  421 LysLysSerLeuGlyLysLysArgLeuArgLeuGlnAlaSerArgLeuGlyGlyValSerSerSerGlnThr  440
          ||||||||||||||||||||||||||||||||||||||||||-----------GGCACC 1499
Db 1457 AAGAAGAGCCTGGGGAAGCGGCGTCTCCGGCTCCAGGCCAGCTG

Qy  441 ThrLysGlnLeuProAlaSerLeuLeuHisProLeuHisGlnAlaMetThrAspProLeu  460
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1499 ACAAAACAACTGCCAGCCTCCTTCTTCATCCACTCTTCAGGCAATGACTGACCCACTG 1558

Qy  461 CysHisLeuSerLeuThrLeuSerHisCysLysLysLeuProAspAlaValCysArgAsp  480
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1559 TGCCATCTGAGCCTCACGCTGTCCCACTGCAAACTCCCGACGCGGTCTGCCGAGAC 1618

Qy  481 LeuSerGluAlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuLeuHisAsnArg  500
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1619 CTTTCTGAGGCCCTGAGGGCAGCAGCCCCGGCACTGACGGAGCTGGGCCTCCTCCACAACAGG 1678

Qy  501 LeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgVal  520
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1679 CTCAGTGAGGCAGGACTGCGATGCTGAGTGAGGGCCTAGCCTGGCCGCAGTGCAGGGTG 1738

Qy  521 GlnThrValArgValGlnLeuProAspProGlnArgGlyValLeuGlnTyrLeuValGlyMet  540
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1739 CAGACGGTCAGGGTACAGTTGCCTGACCCCCAGGAGCGGGGTACCTGGTGGGTATG 1798

Qy  541 LeuArgGlnSerProAlaLeuThrLeuSerGlyCysGlnLeuProAlaPro  560
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1799 CTTCGGCAGAGCCCTGCCCTGACCCTGAGCGGCTGCCAACTGCCCGCCCCC 1858

Qy  561 MetValThrTyrLeuLeuValGluHisGlnGlyValLeuGlnThrLeuSer  580
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1859 ATGGTGACCTACCTGTGCAGTCGTGCAGCACCAGGATCGGCCTGCAGACCCTCAGT 1918

Qy  581 LeuAlaSerValGluLeuSerGluLeuGlnSerLeuGlnLeuGluLeuGlnAlaValLysArgAla  600
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1919 CTGGCCCTGTCGTGGACGCGAGCGAGCTCACTACAGGAGCTTCAGGTTGAGAGAGCA 1978

Qy  601 LysProAspLeuValIleThrHisProAlaLeuAspProArgThrAlaAlaGlyValyVal  620
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db 1979 AAGCCCGGATCTGGTCATCACACAACCCCAGCACCCAACCTCCCAAGGAA 2038
```

```
Qy    621 LeuIleSerThrPhe 625
          |||||||||||||||
Db   2039 CTCATCTCGACCTTC 2051
```

RESULT 5
AAL47132

ID   AAL47132 standard; DNA; 5556 BP.
XX
AC   AAL47132;
XX
DT   20-AUG-2002 (first entry)
XX
DE   Pyrin domain containing protein NALP6/PY9 coding sequence.
XX
KW   Pyrin domain; PYD domain; antiinflammatory; antiparkinsonian;
KW   antiarteriosclerotic; antipsoriatic; antibacterial; virucide;
KW   neuroprotective; antiarthritic; antirheumatic; antiasthmatic;
KW   nephrotropic; osteopathic; nootropic; intracellular signal transduction;
KW   inflammation; Alzheimer's disease; infection; psoriasis; asthma;
KW   arteriosclerosis; multiple sclerosis; rheumatoid arthritis; sarcoidosis;
KW   osteoarthritis; glomerulonephritis; gene; ds.
CS   Unidentified.
XX
PN   WO200240668-A2.
XX
PD   23-MAY-2002.
XX
PF   30-OCT-2001; 2001WO-EP12545.
XX
PR   15-NOV-2000; 2000DE-1056687.
PR   30-NOV-2000; 2000DE-1059595.
XX
PA   (APOT-) APOTECH RES & DEV LTD.
XX
PI   Tschopp J, Martinon F;
XX
DR   WPI; 2002-427093/45.
DR   P-PSDB; AAO17860.
XX
PT   New DNA encoding protein with pyrin domain, useful for treating
PT   diseases involving impaired signal transduction, particularly
PT   inflammation, also proteins and antibodies -
XX
PS   Claim 5; Fig 1; 116pp; German.
XX
CC   The present invention relates the DNA and their encoded proteins, where
CC   the proteins contain at least one PYD (pyrin) domain. These can be used
CC   to treat diseases associated with impaired intracellular signal
CC   transduction, particularly inflammation such as psoriasis,
CC   arteriosclerosis, bacterial or viral infections (particularly meningitis
CC   and pneumonia), multiple sclerosis, rheumatoid arthritis, asthma,
CC   sarcoidosis, glomerulonephritis and osteoarthritis, and also Alzheimer's
CC   and Parkinson's diseases. The present sequence is a coding sequence of
CC   the invention.
XX
SQ   Sequence 5556 BP; 914 A; 1856 C; 1879 G; 907 T; 0 other;

Alignment Scores:
Pred. No.:
Score:                  2882.00
Percent Similarity:      82.92%        Length:         5556
Best Local Similarity:   81.83%        Matches:         581
Query Match:             89.56%        Conservative:      7
DB:                                    Mismatches:       32
                                       Indels:           90
                                       Gaps:              5

US-10-029-347-2  (1-625) x AAL47132 (1-5556)

```
Qy      1 MetLeuAlaGlyProGlnArgProLeuLeuPheLeuLeuAspProCysThrAspProGluAlaAlaProGlyAlaAspGluLeuProAla   20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db    799 ATGCTGGCCGGACCCCAGCGGCCGCTGCTCTTCATCCTGGACCCGCGCGGACCAGCTGCCGCG   858

Qy     21 LeuGlyGlyProGluAlaAlaProCysThrAspProCysThrAspProGluAlaAlaSerGlyAlaAlaArg   40
          |||||||||||||||||||
Db    859 CTGGGGGGCCCCGAGGCCGGCGGCCGGCCCTGCACAGACCCCTGCACAGACCCTCGAGCCGGCGGAGGGGCGGCCGCGGCGG   918
```

```
Qy     41 ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValThrThr   60
          |||||||||||||||||||||||||||||||||||||||||||||
Db    919 GTGCTAGGCGGGCTGCTGAGCAAGGCGCTCCTGCCCACGGCCCTCCTGGTGACCACG   979

Qy     61 ArgAlaAlaAlaProGlyArgLeuGlnLeuArgLeuCysSerProGlnCysAlaGluVal   80
          |||||||||||||||||||||||||||||||||||||||
Db    979 CGGGCGGCCGCCCCCGGGAGGCTGCAGCTGCGCCTGTGTTCCCCGCAGTGCGCCGAGGTG   1038

Qy     81 ArgGlyPheSerAspLysAspLysLysLysTyrPhePheArgAspGluArg   100
          |||||||||||||||||||||||||||||||||||||||
Db   1039 CGCGGCTTCTCGACAAGGACAAGAAGAAGTATTTCTTCCGCGGATGAGAGG   1098

Qy    101 ArgAlaGluArgAlaTyrArgArgPheValGluAsnGluThrLeuAsnGluThrLeuAlaLeuCysPhe   120
          |||||||||||||||||||||||||||||||||||||||
Db   1099 AGGGCCGAGCGCGCCTACCGCCGCTTCGTGGAAGGAGACGAGACGCTGAATGAGACGCTGGCCCTGTGTTTC   1158

Qy    121 ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnIleGluGlnLeuGlyGlyArg   140
          |||||||||||||||||||||||||||||||||||
Db   1159 GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGCTCGGTCGG   1218

Qy    141 AspLeuSerArgThrSerLysThrThrSerValTyrLeuValPheIleThrSerVal   160
          |||||||||||||||||||||||||||||||||||
Db   1219 GACCTGTCGCGCACGTCCAAGACCACCAGTGTACCTGCTTTTCATCACCAGCGTT   1278

Qy    161 LeuSerSerAlaProValAlaAspGlyArgLeuGlnGlyArgLeuArgLeuArgAsnLeuCys   180
          |||||||||||||||||||||||||||||||||||
Db   1279 CTGAGCTCGGCTCCGGTAGCCGACGGGCGCCTGCAGGGCGACCTGCGAATCTGTGC   1338

Qy    181 ArgLeuAlaArgGluGlyValLeuGlyGlyValArgArgAlaGlnPheAlaGlnIleGluGlu   200
          |||||||||||||||||||||||||||||||||||
Db   1339 CGCCTGGCCCGCGGGCGGCGTCCCGGACGCGCAGGGCGCAGTTGCCGAAGAACTGGAG   1398

Qy    201 GlnLeuGluLeuArgGlySerLysSerValGlnThrLeuPheLeuSerLysLysSerGluLeuPro   220
          |||||||||||||||||||||||||||||||||||
Db   1399 CAACTGGAGCTTCGTGGCTCCAAAGTGCAGACGCTGTTTCTGAGCAAAAAGGAGCTGCCG   1458

Qy    221 GlyValLeuGluThrGluValThrTyrGlnPheLeuGlnSerPheGlnSerPheGlnLeuGluPheLeu   240
          |||||||||||||||||||||||||||||||||||
Db   1459 GGGGTGCTGGAGACAGAGGTCACCTACCAGTTCCTCCAGAGTTCCTC   1518

Qy    241 AlaAlaLeuSerTyrLeuGluLeuTyrProArgThrAlaAlaGlyValGlyVal   260
          |||||||||||||||||||||||||||||||||||
Db   1519 GCGGCACTGTCCTACCTGCTGGAGCACGGCGGCCCCAGGACCGGCGGCGGGCGGTT   1578

Qy    261 GlyThrLeuLeuArgGlyAspAlaGlnProHisSerHisLeuValLeuThrThrArgPhe   280
          |||||||||||||||||||||||||||||||||||
Db   1579 GGGACACTCCTGCGTGGGGACGCGCAGCCCCACAGCCACTTGGTGCTCACCACGCGGCTTC   1638

Qy    281 LeuPheGlyLeuLeuSerAlaGlyArgMetArgAspIleGlyArgHisPheGlyCysMet   300
          |||||||||||||||||||||||||||||||||||
Db   1639 CTCTTCGGACTGCTGAGCGCCGGACGCGGCCGCATGCGAGACATCGGCCGCCACTTCGGCTGCATG   1698

Qy    301 ValSerGluArgValAlaValGlnGluValAlaLeuArgTrpValGlnGlyGlnGlyGlyCys   320
          |||||||||||||||||||||||||||||||||||
Db   1699 GTTTCAGAGCGTGTGAGCGTGGGGGTGCAGGAGGTGGCTGGCGCGCTGC   1756

Qy    321 ProGlyValAlaProGluValThrGluGlyValLysValLeuAspThrGluLeuPro   340
          |||||||||||||||||||||||||||||||||||
Db   1759 CCCGGAGTGGCACCAGATGAGCTGGAGGGGCTCAAAGGGCTCGAGGACACCGAAGAGCCA   1818

Qy    341 GluGluGluGlyGluGlyGlnProAsnTyrProLeuGluLeuGlyLeuTyrCysLeuTyr   360
          |||||||||||||||||||||||||||||||||||
Db   1819 GAGGAGGAGGGGGAGGAGGAGCCCAACTACCCCCTGGAGTTGTCTACTGCCTCTAC   1878

Qy    361 GluThrGlnGluGluAlaProAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu   380
          |||||||||||||||||||||||||||||||||||
Db   1879 GAGACGCAGGAGGACCGGTTTGTGCGCCAAGCCCTGTCCCGGACCTGGACCTGGCCCTG   1938

Qy    381 GlnArgValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCys   400
          |||||||||||||||||||||||||||||||||||
```

Db  1939  CAGCGGAGTGCGCTTCTGCGCATGGACGTGGCTGTTCTGAGCTACTGCGTGAGGTGTGC  1998

Qy  401  ProAlaGlyGlnAlaLeuArgLeuLeuSerCysArgCysLeuValAlaAlaGlnGluLysLys  420
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1999  CCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGGAGAAGAAG  2058

Qy  421  LysLysSerLeuLysArgLeuGlnAlaSerLeuGlyGlySerGlnGlyThr  440
         ||||||||||||||||||||||||||||||||||||||||||||||||||
Db  2059  AAGAGAGCTGGGGAAGCGGCTTCAGGCAGCCTGGCGTGGCCGGCAGCTGGCTGGGGACC  2118

Qy  440  ----  440

Db  2119  CAACTGGCTCCAGAAGTACCCTTTCGACCACCCTGCTGTGACATCTGCCCACACCTCCA  2178

Qy  440  ----  440

Db  2179  CCAGAACCTCCGGCTCCTCCAGGGCAAGGCTTTTGCCAGAGTTCCTTTGAAATTAGCCCA  2238

Qy  441  ThrLysGlnLeuProAlaSerLeuLeu  449
         :::    |||||
Db  2239  ATTCAGCCCCTGCCCAGGGGCTTGGCATCTGTTGAGAGGATGAAGTCACGGTGTTGGCA  2298

Qy  450  ------HisProLeuPheGlnAlaMetThrArgProLeuCysHisLeuSer  464
              ||||||||||||||||||||
Db  2299  GGGGCTGGGCCTGGGGACCCAAAGACCCCATGCAATGACCTGACCCACCTGTGCCATCTGAGC  2358

Qy  465  SerLeuThrLeuSerHisCysLysLeuProAlaValCysArgArgAspLeuSerGluAla  484
         ||||||||||||||||||||||||||||||||
Db  2359  AGCCTCACGCTGTCCCACTGCAAACTCCCCGACGCGGTCTGCCGGAGACCTTTCTGAGGCC  2418

Qy  485  LeuArgAlaAlaProAlaLeuThrGlnLeuGlyLeuLeuHisAsnArgLeuSerGluAla  504
         |||||||||||||||||||||||
Db  2419  CTGAGGGCAGCCCCGCACTGACGGAGCTGGGCCTCCTCCACAACACAGGCTCAGTGAGGCA  2478

Qy  505  GlyLeuArgMetLeuSerGluGluGlyLeuLeuAlaTrpProGlnCysArgValGlnThrValArg  524
         ||||||||||||||||||||||||||||
Db  2479  GGACTGCGTATGCTGAGTGAGGGCCTGGCCGTGCAGGGTGCAGACGGTCAGG  2538

Qy  525  ValGlnLeuProAspProGlnArgGlyArgGlnTyrLeuValGlyMetLeuArgGlnSer  544
         :::::::::::
Db  2539  GTACAGCTGCCTGACCCCCAGCGGGGTCCAGTACCTGGTGGGTATGCTCGGCAGAGC  2596

Qy  545  ProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyr  564
         ||||||||||||||||||||
Db  2599  CCTGCCCTGACCACCCTGGATCTCTCCGGATGCCAACTGCCCGCCCCCATGGTGACCTAC  2658

Qy  565  LeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSerLeuAla------  582
         ||||||||
Db  2659  CTGTGTGCAGTCCTGCAGCACCAGGGATGCGGCCTGCAGACCCTCAGCCTCTCGCTTCCT  2718

Qy  583  ------------------SerValGluLeuSerGlu-----------  588
         |||
Db  2719  TCTGACCCGACCCCGAGTTC-TTCTCCGGACGGTGTCGAGAACCCGGGCCCGGCTGGGG  2778

Qy  589  ------------GlnSerLeuGlnLeuGluLeuGlnGlnAlaValLysArgAlaLysPro  602
         :::::||||||
Db  2779  CTGAGTCTTCGCTCGGCAGCGCCCCCCGAGCCCCTGGGCACAGCGAGGCGGAGGACC  2838

Qy  603  AspLeuValIleThr-------------------HisProAla  610
         ||||||||
Db  2839  CAGGTGGAGGGCGGCGGCGGGGGGCGCGGGGAGGGAGGCGCGGGAGGGAACCCCCGGCC  2898

Qy  611  LeuAspGlyHisProGlnProProLysGlu  620
         :::::::::
Db  2899  CGGCGGGCCCCACCCACAGCCGCCCCCGGGAC  2928

RESULT 6
AAS00033
ID   AAS00033 standard; cDNA; 6461 BP.
XX
AC   AAS00033;
XX
DT   10-MAY-2001 (first entry)

XX
DE   Human ATLAS-2-encoding cDNA.
XX
KW   Human; Activated T-lymphocyte associated sequence 2; ATLAS-2; antibody;
KW   cytokine receptor; autoimmune disorder; immune disorder; cancer;
KW   T-lymphocyte-associated disorder; cell-proliferation disorder; tumour;
KW   cell differentiation disorder; immune deficiency disorder; malignancy;
KW   viral infection; bacterial infection; fungal infection; metabolism;
KW   chromosome 11p15.5; ss.
XX
OS   Homo sapiens.
XX
FH   Key          Location/Qualifiers
FT   CDS          1..5556
FT                /tag= a
FT                /product= "ATLAS-2"
XX
PN   WO200114564-A2.
XX
PD   01-MAR-2001.
XX
PF   18-AUG-2000; 2000WO-US22699.
XX
PR   20-AUG-1999;  99US-0150105.
PR   28-APR-2000; 2000US-0560101.
PR   28-APR-2000; 2000US-0560365.
PR   28-APR-2000; 2000US-0560948.
PR   28-APR-2000; 2000US-0561533.
XX
PA   (CURA-) CURAGEN CORP.
PA   (BIOC ) BIOGEN INC.
XX
PI   Feyman JA, Green CD, Hsu A, Browning JA, Carulli J;
XX
DR   WPI; 2001-218453/22.
DR   P-PSDB; AAU00023.
XX
PT   New isolated activated T lymphocyte associated sequences for treating
PT   or preventing immune system associated disorders such as autoimmune
PT   disorder, immune disorder, and T-lymphocyte-associated disorder -
XX
PS   Claim 6; Fig 2; 114pp; English.
XX
CC   The sequence encodes human activated T-lymphocyte associated sequence 2,
CC   ATLAS-2. ATLAS-2 is related by homology to cytokine receptors and its
CC   gene is located on chromosome 11p15.5. ATLAS proteins, polynucleotides
CC   and antibodies are useful for treating/preventing conditions associated
CC   with an autoimmune disorder, immune disorder, T-lymphocyte-associated
CC   disorder, cell-proliferation disorder, cell differentiation disorder,
CC   and immune deficiency disorder and for screening for a modulator of
CC   activity or of latency or predisposition to an immune disorder.
CC   ATLAS proteins, polynucleotides and antibodies are useful in therapeutic
CC   or prophylactic treatment of diseases associated with cell-proliferation
CC   (e.g. cancers, malignancies and tumours). The polynucleotides are useful
CC   in gene therapy, to detect ATLAS mRNA or a genetic lesion in an ATLAS
CC   gene, to modulate ATLAS activity, to screen drugs or compounds that
CC   modulate ATLAS activity or expression and to treat disorders
CC   characterised by insufficient or excessive production of ATLAS protein or
CC   production of ATLAS protein forms that have decreased or aberrant
CC   activity compared to ATLAS wild type protein and in tissue typing to
CC   identify individuals. The antibodies are useful for localisation/
CC   quantitation, isolation and detection of ATLAS and to monitor protein
CC   levels in tissue. ATLAS is useful for treating/preventing infection by
CC   bacteria, viruses and fungi, affecting bodily characteristics, e.g.
CC   biorhythms, fertility or metabolism, affecting behavioural
CC   characteristics, and for providing analgesic effects. A host cell
CC   containing the polynucleotide is useful to produce non-human transgenic
CC   animals.
XX
SQ   Sequence 6461 BP; 1079 A; 2146 C; 2150 G; 1086 T; 0 other;

Alignment Scores:
Pred. No.:        6.19e-240        Length:        6461

```
Score:                    2882.00         Matches:        58?
Percent Similarity:       82.82%          Conservative:    7
Best Local Similarity:    81.83%          Mismatches:     32
Query Match:              89.56%          Indels:         90
DB:                                       Gaps:            5

US-10-029-347-2 (1-625) x AAS00033 (1-6461)
```

```
Qy      1  MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla    20
           |  ||| |||| |||||| ||||||| ||||||||||||||||||| ||||||
Db    799  ATGTGGCCCCAGCCGCCGCTCTCTTCATCCTGGACGGCCGGCGAGCTGCCCGCG    855

Qy     21  LeuLeuGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg    40
           ||| ||||||||||||||||||||| ||||||||||||||||||||||||||
Db    859  CTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCG    918

Qy     41  ValLeuGlyGlyLeuLeuSerLysAlaLeuArgLeuProThrAlaLeuLeuValThrThr    60
           |||||||||| |||||||||||||||||||||||||||||||||||||||||
Db    919  GTGCTAGGCGGGCTGCTGAGCAAGGCGGTGCTGCGGCTGCCCACGGCCCTGCTGGTGACCACG    978

Qy     61  ArgAlaAlaAlaProGlyGlnGlyArgLeuCysSerProGlnCysAlaGluVal    80
           |||||||||| |||||||||||||||||||||||||||||||||||||
Db    979  CGGGCGGCGGCGCCCGGGCAGGGCCGCCTGTGTTCCCCGCAGTGCGCCGAGGTG   1038

Qy     81  ArgGlyPheSerAspLysLysLysTyrPheTyrLysPhePheArgArgGluArg   100
           ||||| ||||||| ||||||||||||||||||||||||||||||||||||
Db   1039  CGCGGCTTCTCCGACAAGAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGAGG   1098

Qy    101  ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe   120
           ||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   1099  CGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC   1158

Qy    121  ValProPheValCysTrpIleValCysThrLeuArgValLeuGlnLeuLeuGlyArg   140
           |||||| ||| |||||||| ||||||||||||||||||||||||||
Db   1159  GTGCCCTTCGTGTGCTGGATCGTGTGCACCCTGCGCGTGCTGCAGCTGCTGGGCCGG   1218

Qy    141  AspLeuSerArgThrSerLysThrThrArgValTyrLeuLeuPheIleThrSerVal   160
           ||||| ||||||||||| |||||||||||||||||||||||||||
Db   1219  GACCTGTCGCGCACGTCCAAGACCACCAGGGTCTACCTGCTTTTCATCACCAGCGTT   1278

Qy    161  LeuSerSerAlaProValAlaAspGlyTyrProArgLeuArgAsnLeuCys   180
           ||||| |||||||||| |||||||||||||||||||||||||||||
Db   1279  CTGAGCTCGGCTCCGGTAGCCGACGGGCCCCGGCTGCGGAACCTGTGC   1338

Qy    181  ArgLeuAlaAlaArgGlyGluGlyValAlaValLysAlaGlnPheAlaGlyLysGlyLeuGlu   200
           |||||||||||||| ||||||||||| ||||||||||||||||||||||||||
Db   1339  CGCCTGGCCGCGCGAGGCGAGGGCGTCGCCGTGAAGGCGCAGTTTGCCGAGAAGAACTGGAG   1398

Qy    201  GlnLeuGluArgGlySerValGlnThrLeuPheLeuSerLysLysGluProPro   220
           |||||| |||||||||| ||||||||||||||||||||||||||
Db   1399  CAACTGGAGCGGCTCGGTCCAAAGTGCAGACGCTGTTTCTCAGCAAAAAGGAGCCGCCG   1458

Qy    221  GlyValIleLeuGluThrGlnPheGlnPheIleAspSerPheGlnLeuHisLeu   240
           ||||| |||||||||||||||||| |||||||||||||||||
Db   1459  GGGGTGATCCTGGAGACGCAGTTCCAGTTCATCGACAGCTTCCAGCTTCCTC   1518

Qy    241  AlaAlaLeuSerTyrLeuLeuGluAspTyrLysValProArgAlaGlyValAlaGlyVal   260
           |||||||||| |||||||||||||| ||||||||||||||||||||
Db   1519  GCGGCACTGTCCTACCTGCTGGAGGACGGCGGGGTGCCCCGGGCCGGCGTGGCGGGCGT   1578

Qy    261  GlyThrLeuArgGlyLeuLeuSerAlaGluLeuArgMetArgAspIleGluLeuValLeuThrThrArgPhe   280
           ||||||| |||||||||||||||| |||||||||||||||||||||||||||||||
Db   1579  GGGACACTGCTGAGCGCCGAGCTGAGGATGCGGGACATCGAGCTGGTGCTCACCACGCGGTTC   1638

Qy    281  LeuPheGlyLeuLeuSerAlaGluValGlyMetArgAspIleGlyHisLeuSerPheCysMet   300
           ||||| |||||| |||||||||| ||||||| ||||||||||||||||||||||||
Db   1639  CTCTTCGGACTGCTGAGCGCCGAGGTGGGCATGCGGGACATCGGGCACTTCGGCTGCATG   1698

Qy    301  ValSerArgValLysGlyGlnAlaLeuAlaLeuArgTrpValTrpIleGlyIleGlyGlyCys   320
           ||||| ||||||||||||| ||||||||||||||||||||||||||||||||
Db   1699  GTTTCAGCGGTGTGAAGGGCCAGGCGCTGGCGTTGCGGTGGGTGTGGATTGGCATCGGTGGCTGC   1758

Qy    321  ProGlyValAlaProGluThrGlyAlaLysGlyAlaLysGluAspThrGlu...   340
           ||| |||||||||
Db   1759  CCCGGAGTGGCACCAGAGGTGACCGGAGGGGCCAAGGGCTCGGAGGACACCGAAGAGCCA   1818

Qy    341  GluThrGlnGluGluGlyGluGluGluGlnProAsnTyrProLeuGluLeuTyrCysLeuTyr   360
           ||||||||||||||||||||||||||||||||||||||||||||||
Db   1819  GAGAGGAGGAGGAGGAGGAGGAGCCAACTACCCACTGGAGTTGCTACTGCCTGTAC   1878

Qy    361  GluThrGlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeu   380
           |||| |||||| ||||| |||||||||||||||||||||||||||||||||
Db   1879  GAGACGCAGGAGGACGCGGTTTGTGCGCAGGCCCTGTGCCGGTTCCCGGAGCTGGCGCTG   1938

Qy    381  GlnArgValPheCysArgMetAspAspValAlaValLeuSerTyrCysValArgCysCys   400
           ||||||||| |||||||||||||||||||||||||||||||||||||||
Db   1939  CAGCGAGTGCCGCTTCTGCCCCATGGACGACGTGGCTGTTCTGAGCTACTGCGTGAGGTGCTGC   1998

Qy    401  ProAlaGlyAsnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaLeuGlnLeuLysLys   420
           ||||||||| |||||||||||||| |||||||||||||||||||||||||
Db   1999  CCTGCTGGACAGGCACTGCGGCTGATCAGCTGCAGATTGGTTGCTGCGCAGGAGAAGAAG   2058

Qy    421  LysLysSerLeuGlyLysArgGlnLeuAlaSerLeuGlyGlySerGlnGlyThr   440
           ||||||||||||||| |||||||||||||||||||||||||||||
Db   2059  AAGAAGAGCCTGGGAAGCCGGCTCGCCAGGCCAGCCTGGGCTGGCTGGGACC   2118

Qy    440  ...   440
           
Db   2119  CAACTGGCTCCAGAGTACCCTTTCGACACCCTGCGTGTGACATCTGCCCACACCTCCA   2178

Qy    440  ...   440
           
Db   2179  CCAGACCCTCCGTGCCTCCTCCAGGCAAGGCGTTTGCCAGAGTGAATATAGCTCCA   2238

Qy    441  ThrLysGlnLeuProAlaSerLeuLeu...   449
           ::::
Db   2239  ATTCAGCCCCTGCCCAGGGGCTTGCATCTGTTGAGGAGGATGAATGTCACGGTGTTGGCA   2298

Qy    450  ...HisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSer   464
                  ------------||||||||||||||||||||||||
Db   2299  GGGGGTGGGCCTGGGCACCAAAGACCCATGCAATGACTGACCCACTGTGCCATCTGAGC   2358

Qy    465  SerLeuThrLeuSerHisCysLysLeuProAlaSerAlaCysArgAspLeuSerGluAla   484
           ||||||||||||||| |||||||||||||||||||||||||||||||||||
Db   2359  AGCCTCACGTCTGTCCCACTGCCAAACTCCCTGCCGCAGCGGTCTGCCCAGACCTTTCTGAGGCC   2418

Qy    485  LeuArgAlaAlaProAlaLeuThrGlyLeuLeuLeuHisAsnArgLeuSerGluAla   504
           |||||||||||||||||||||||| |||||||||||||||||||||||||
Db   2419  CTGAGGGCAGCGCCCCCACTCACGGAGCTGGGCCTTCCTCCACAACAGGCTCAGTGAGGCA   2478

Qy    505  GlyLeuArgMetLeuSerGluGlyValAlaTrpProGlnCysArgValGlnThrValArg   524
           ||||||||| |||||||||||| |||||||||||||||||||||||
Db   2479  GGACTGCGTATGCTGAGTGAGGGCGTCAGTGCAGGGTGCAGACGGTCAGG   2538

Qy    525  ValGlnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSer   544
           |||||||||||||||||||||||||||||||||||||||||||||
Db   2539  GTACAGCTGCCTGACCCCCAGGAGCCCCGAGGGCTCCAGTACCTGGTGGGTATGCTTCGGCAGAGC   2598

Qy    545  ProAlaLeuThrLeuAspLeuSerGlnCysGlnLeuAspLeuSerLeuAla...   564
           ||||||||||||||||||||||||||||||||||||||||||||||||
Db   2599  CCTGCCCTGACACTGGACCTCAGCCAGTGTCAGCTGGATCTCAGCCTGGCACTGGTGACCTAC   2658

Qy    565  LeuCysAlaValLeuGlnHisGlnHisCysLysGlyLeuGlnThrLeuSerLeuAla...   582
                                       ------SerValGluLeuSerGlu
Db   2659  CTGTGTGCAGCCTCTGACCCCGAGTTCTTCTCCGGACACCCGGCCTGGGCCGGCTGGGG   2718

Qy    583  ...SerValGluLeuSerGlu...   588
           ||||||||||
Db   2719  TCTGACGGCAGCCTGTGTGCAGCCTTCTTCTCCGGAGACACCCGGCCTGGGCCGGCTGGGG   2778

Qy    589  ...GlnSerLeuGlnGluLeuGlnHisAlaVaLysArgAlaAlaAlaLysPro   602
                  ------:::||
Db   2779  CTGGAGTCTCGCTGGCCTGGGACGCCCTCGGGGCGACAGCGAGGACGAGGACC   2838

Qy    603  AspLeuValIleThr...   610
           :::
Db           ------HisProAla   610
```

```
Db   2839  CAGGTGGAGCGCGCGGGGCGCGCGGGCGGAGGCGCGGAGGAACCCCGGCC 2898
Qy    611  LeuAspGlyHisProGlnProProLysGlu 620
                 ||||||||||||||||||::::::
Db   2899  CGCGGCGCCCACCCAGCCGCCCCGGGAC 2928

RESULT 7
ABA08979
ID   ABA08979 standard; cDNA; 1795 BP.
XX
AC   ABA08979;
XX
DT   11-JAN-2002 (first entry)
XX
DE   Human vasopressin receptor homologue-encoding cDNA, SEQ ID NO:755.
XX
KW   Human; cytokine; cell proliferation; cell differentiation; growth factor;
KW   haematopoiesis regulation; tissue growth; immunomodulator; activin;
KW   inhibin; chemotaxis; chemokinesis; thrombolysis; oncogenesis;
KW   proliferation; metastasis; cancer; tumour; haematopoietic disorder;
KW   myeloid cell disorder; lymphoid cell disorder; asthma; arthritis;
KW   chronic inflammatory condition; proliferative retinopathy;
KW   atherosclerosis; coronary heart disease; arterial ischaemia;
KW   bone disorder; osteoporosis; vascular growth disorder;
KW   tissue regeneration; wound healing; infection; immune disorder;
KW   cell culture; drug screening; gene therapy; antiinflammatory;
KW   antiasthmatic; antiarthritic; haemostatic; antiarteriosclerotic;
KW   cytostatic; osteopathic; vasotropic; cardiant; virucide; antibacterial;
KW   antifungal; vulnerary; antiulcer; ss.
XX
CS   Homo sapiens.
XX
PN   WO200157162-A2.
XX
PD   09-AUG-2001.
XX
PF   05-FEB-2001; 2001WO-US03800.
XX
PR   03-FEB-2000; 2000US-0496914.
PR   27-APR-2000; 2000US-0560875.
XX
PA   (HYSE-) HYSEQ INC.
XX
PI   Tang YT,  Liu C,  Drmanac RT;
XX
DR   WPI; 2001-457740/49.
DR   P-PSDB; ABB11735.
XX
PT   Human proteins and DNA encoding sequences useful for preventing,
PT   treating or ameliorating a medical condition in a mammalian subject
PT   e.g. arthritis and cancer -
XX
PS   Claim 1; Page 678-679; 1963pp; English.
XX
CC   Sequences ABB10981-ABB12330 represent 1350 novel human polypeptides, and
CC   sequences ABA08225-ABA09574 represent nucleic acids encoding them. The
CC   invention also relates to vectors and recombinant host cells comprising a
CC   nucleotide of the invention, methods of producing the novel polypeptides,
CC   antibodies against the polypeptides, methods of detecting the nucleotides
CC   or polypeptides in a sample, and methods of identifying compounds which
CC   bind to polypeptides of the invention. Although novel, many of the
CC   polypeptides of the invention have homology to known proteins, thereby
CC   giving an insight into their probable biological activities, and hence
CC   potential therapeutic applications. The polypeptides of the invention may
CC   have various activities, including cytokine, cell proliferation or cell
CC   differentiation activities; stem cell growth factor activity;
CC   haematopoiesis regulatory activity; tissue growth activity;
CC   immunomodulatory activity; activin- or inhibin-related activities;
CC   chemotactic or chemokinetic activities; haemostatic, thrombotic or
CC   thrombolytic activities; receptor or ligand activities; or may be
CC   involved in oncogenesis, cancer cell proliferation or metastasis.
CC   Depending on their biological activities, polypeptides and nucleotides of
CC   the invention are useful for preventing, treating or ameliorating medica-
```

```
CC   conditions, e.g., by protein or gene therapy. Such conditions include
CC   cancers, haematopoietic disorders (e.g.,myeloid or lymphoid cell
CC   disorders), chronic inflammatory conditions (e.g., asthma or arthritis),
CC   proliferative retinopathy, atherosclerosis, coronary heart disease,
CC   arterial ischaemia, bone disorders (e.g., osteoporosis), and abnormal
CC   vascular growth. Polypeptides involved with tissue regeneration and
CC   repair (or nucleic acids encoding them) may be used to promote wound
CC   healing (e.g., of burns, incisions and ulcers), while those with
CC   immunomodulatory activities may be used in the treatment of viral,
CC   bacterial and fungal infections in addition to immune disorders.
CC   Polypeptides with growth factor activity may be used in cell cultures to
CC   promote cell growth. For example, such polypeptides may be used to
CC   manipulate stem cells in culture to give rise to neuroepithelial cells
CC   that can be used to augment or replace cells damaged by illness,
CC   autoimmune disease or accidental damage. The polypeptides and nucleotides
CC   may also be used in the diagnosis of the above conditions, and in drug
CC   screening techniques. The present sequence represents a cDNA encoding a
XX   novel human polypeptide of the invention.
SQ   Sequence 1795 BP; 297 A; 567 C; 656 G; 275 T; 0 other;

Alignment Scores:
Pred. No.:                              Length:         1795
Score:               8.41e-185          Matches:        435
Percent Similarity:  2239.00            Conservative:   0
Best Local Similarity: 100.00%          Mismatches:     0
Query Match:         100.00%            Indels:         0
DB:                  69.58%             Gaps:           0
                     22

US-10-029-347-2 (1-625) x ABA08979 (1-1795)

Qy     1  MetLeuAlaGlnProGlnArgLeuLeuPheLeuAspGlyAlaAspGluLeuProAla  20
Db   489  ATGCTGGCCCAGCCGCAGCGGCTGCTTCTCATCCTGGACGGCGCGGACGAGCTGCCGGCG 548

Qy    21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg  40
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   549  CTGGGGGGCCCCGAGGCGGCGCCCTGCACAGACCCCTTCGAGGCGGCCAGCGGCGCCCGG 608

Qy    41  ValLeuGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr  60
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   609  GTGCTAGGGCGGCTGCTGAGCAAGGCGCTGCTGCCCACCGCCCTCCTGCTGGTGACCACG 668

Qy    61  ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysSerAlaGlyVal  80
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   669  CGGCGCGCCGCGCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCGGCAGTGCGCGGAGGTG 728

Qy    81  ArgGlyPheSerAspLysAspLysAspGlyPheTyrLysPheHisAspArgAspGluArg  100
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   729  CGGGGCTTCTCCGACAAGGACAAGGACGGATTCTACAAGTTCCACGATCGAGAGG 788

Qy   101  ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrArgLeuAlaLeuCysPhe  120
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   789  AGGGCCGAGCGCGCGCTACCGCTTCGTGAAGGAGAACGAGACGCGTCTTCGGCTGTGCTTC 848

Qy   121  ValProPheSerValCysTrpIleValCysThrValLeuArgGlnGlnLeuGlyLeuGlyArg  140
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   849  GTGCCCTTCTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGGTCGCGG 908

Qy   141  AspLeuSerArgThrSerLysThrThrSerProValTyrLeuLeuPheLeuThrSerVal  160
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   909  GACCTGTCGCGCACGTCCAAGACCACCAGTACGTACCCTGTGTTTCATCACCAGCGTT 968

Qy   161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlyGlyAspLeuArgAsnLeuCys  180
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   969  CTGAGCTCGGCCCCGGTAGCCGACGGCCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGC 1028

Qy   181  ArgLeuAlaArgGluGlyValLeuGlyValArgArgAlaGlnPheAlaGluLysGluLeuGlu  200
               ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  1029  CGCCTGGCCCGCGAGGGCGTCCTCCGACGCAGCCAGTTTGCCGAGAAGAACTGGAG 1088

Qy   201  GlnLeuGluLeuArgGlyLySerLysGlyValGlnThrLeuPheLeuLysSerLysLysGluLeuPro  220
```

Qy   521   GlnThrValArgValGlnLeuProGlnProGlnArgGlyLeuGlnTyrLeuValGlyMet   540
            |||||||||||||| ||||| :||||||||||||| |||
Db   1240   CAGACACTCAGGATACAGATGCCTGGGCTCCAAGAGGTGATCCATTACCGGTCATTGTG   1299

Qy   541   LeuArgGlnSerProAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaPro   560
            ||||| ::|||||
Db   1300   CTCCAGCAGAGCCCAGTCCTAACCACTCTGGACCTGTCCAGCTGCCTGCCTGGGACT   1359

Qy   561   MetValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysGlyLeuGlnThrLeuSer   580
            :||| |||| || ::|||| |||||
Db   1360   GTGGTGGAACCTCGTGTTCAGCCTGAAACACCCTAAATGTGGCTAAAGACCCTCAGT   1419

Qy   581   LeuAlaSerValGlyGluGlnSerLeuGlnLeuGlnAlaValLysArgAla   600
            |||:|| ||||||
Db   1420   CTGACTTCTGTGGAGCTGACTGGACTGGAGAGAGCTTCAAGCTGTGAAGACATTA   1479

Qy   601   LysProAspLeuValIleThrHisProAlaLeuAspGlnHisProGlnProLysGlu   620
            ||||
Db   1480   AAGCCAGATCTGGCCATCATACATTCAAAATTGGGCACACATCCCTCAGCCTCTGAAGGGA   1539

Qy   621   LeuIleSerThrPhe   625
            ||||
Db   1540   TGAAGCAGTGTCTTC   1554

RESULT 10
AAK83313
ID   AAK83313   standard; DNA; 933 BP.
XX
AC   AAK83313;
XX
DT   07-NOV-2001 (first entry)
XX
DE   Human immune/haematopoietic antigen genomic sequence SEQ ID NO:38125.
XX
KW   Human; immune; haematopoietic; immune/haematopoietic antigen; cancer;
KW   cytostatic; gene therapy; vaccine; metastasis; ds.
XX
OS   Homo sapiens.
XX
PN   WO200157182-A2.
XX
PD   09-AUG-2001.
XX
PF   17-JAN-2001; 2001WO-US01354.
XX
PR   31-JAN-2000; 2000US-0179965.
PR   04-FEB-2000; 2000US-0180628.
PR   24-FEB-2000; 2000US-0184464.
PR   02-MAR-2000; 2000US-0186350.
PR   16-MAR-2000; 2000US-0189874.
PR   17-MAR-2000; 2000US-0190076.
PR   18-APR-2000; 2000US-0198123.
PR   19-MAY-2000; 2000US-0205515.
PR   07-JUN-2000; 2000US-0209467.
PR   28-JUN-2000; 2000US-0214886.
PR   30-JUN-2000; 2000US-0215135.
PR   07-JUL-2000; 2000US-0216647.
PR   07-JUL-2000; 2000US-0216680.
PR   11-JUL-2000; 2000US-0217487.
PR   11-JUL-2000; 2000US-0217496.
PR   14-JUL-2000; 2000US-0218290.
PR   26-JUL-2000; 2000US-0220963.
PR   26-JUL-2000; 2000US-0220964.
PR   14-AUG-2000; 2000US-0224518.
PR   14-AUG-2000; 2000US-0224519.
PR   14-AUG-2000; 2000US-0225213.
PR   14-AUG-2000; 2000US-0225214.
PR   14-AUG-2000; 2000US-0225266.
PR   14-AUG-2000; 2000US-0225267.
PR   14-AUG-2000; 2000US-0225268.
PR   14-AUG-2000; 2000US-0225270.
PR   14-AUG-2000; 2000US-0225447.
PR   14-AUG-2000; 2000US-0225757.

Qy   166   ValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGln   185
            ::|||| :::||||| |||||
Db   187   --ACCAATGACCCCGGGTTCAGGGAGCTGCGAATGCTGTGCCGGCTGGCCCGGCAG   243

Qy   186   GlyValLeuGlyArgArgAlaGlnPheAlaGlnLysGluLeuGluGluLeuArg   205
            |||||| :::|||
Db   244   GGCATCCTGAAGCACATCAAGCACAGTTCTCAGAAAAGGAACCTGGAGAGATTGAAGCTTCAG   303

Qy   206   GlySerLysValGlnThrLeuPheLeuSerLysLysLysGluLeuProGlyValLeuGluThr   225
            ::||||| ||||||
Db   304   GGTTCCCAAGTTCAGACAATGTTTCTCAGCAAGAACCAGCTGCCAGGAGTGCTAGAAACT   363

Qy   226   GlyValThrTyrGlnPheIleAspGlnSerPheGlnHisLeuProLeuAlaAlaLeuSerTyr   245
            |||||
Db   364   GTGGTCACCCACCAGTTCATTGACCAGAGCTTCCAGGAGTTCTGGCTGCATTGTCATAC   423

Qy   246   LeuLeuGluAspGlyGlyValProArgThrAlaAlaGlyValThrLeuGlyThrLeuLeuArg   265
            ||||||
Db   424   CTACTAGAGCGCTGAGGGAGCCCCAGGAACTCCGCAGGAAGTGTGCAGATGCTCCTGCTC   483

Qy   266   GlyAspAlaGlnProHisSerHisValLeuThrThrArgPheLeuPheGlyLeuLeu   285
            ||||||
Db   484   TCTGACGCGGGGGTCCATCTGGTCATCTGGCACTCACCACTAGATTCCTCTTTGGACTGCTA   543

Qy   286   SerAlaGluArgMetArgAspIleGluArgHisPheGlyCysMetValSerGluValArgVal   305
            ||||||
Db   544   AGTACAGAGAGGATTCGTGACATTGGAAACCATTTTGGCTGTGTGGTGTCAGAGGGCGTCTG   603

Qy   306   LysGlnGluAlaLeuArgTrpValGluGlyArgGlyGlnGlyCysGlyTyrVal-Ala---   324
            ::|||||
Db   604   AAACAGGAACACCTTGCGGTGGGTACAAGGCCAAAGCCAA------CCCAAGGTGGCGACA   657

Qy   325   -----ProGluValThrGluGluAlaLysGlyLeuGluAspThrGluLeuProGluGlu   342
            ||||
Db   658   GTAGGAGCAGAAAAGAAGATGAGCTGAAGGACGAGGAGGACGAGGAGGAGGAGGAG   717

Qy   343   GluGluGluGlyGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyrGluThr   362
            |||||
Db   718   GAAGAAGAAGGAGAGGAACTCAACTTTGGACTGGAGCTGTACTGCCTGATGAGACA   777

Qy   363   GlnGluAspAlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArg   382
            ||||| ||||
Db   778   CAAGGAGGATGATTTGTTCGCCAGGCAGGCTCTCAGCAGCC--CCAGAGATGGTACTGGACCGA   837

Qy   383   ValArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysCysProAla   402
            :|||||
Db   838   GTTAGGCTGACCCGCATGGACGTTGAGGTCTGAGCTACTGAGGTGCAGTTGCTGCCCGGAC   897

Qy   403   GlyGlnAlaLeuArgLeuLeuSerGluCysLysLeuThrGluLysAlaAlaGlnGlu-----   420
            |||||
Db   898   GGCCAGGCTCTGAGACTGGTGAGCTGTGGCCGGCAAAGGAGGAGAAGAAGAGAAG   957

Qy   421   LysLysSerLeuGlyLysArgLeuGlnAlaSerArgLeuGlyGlyGlySerSerGlnGlyThr   440
            ||||| ||||
Db   958   AAGAAGAGCTTCATGAACCGGCTGAAG------------GGTTCTCAAAGCACC   999

Qy   441   ThrLysGlnLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrArgProLeu   460
            ||||||
Db   1000   GGGAAACAACCCCCAGCCCTCCTTGCGTGGTCCACTCTGTGAGGCAATGATTACCCAGCAA   1059

Qy   461   CysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAsp   480
            :||||
Db   1060   TGTCATCTGAGTATTCTGACCCTTGCCACACTGCACAGACTGCAGTTGTCGAGAC   1119

Qy   481   LeuSerGluAlaLeuArgAlaLeuProAlaLeuThrGluLeuGlyLeuHisAsnArg   500
            ||||| ||||||
Db   1120   CTTTCTGAGGCTCTGAGGGCTAGCTCCTTCCCTAAGGAGCTGGGCCTCCTCCAGAACCGA   1179

Qy   501   LeuSerGluAlaGlyLeuArgMetLeuSerGluGluLeuAlaTrpProGlnCysArgVal   520
            ||||||
Db   1180   CTCACTGAGGCGAGGGCTGCGTTACTAGAAGCAAGGAAGCCTGGCTTGCCCAAATGCAAGGTG   1239

```
PR   14-AUG-2000;   2000US-0225758.
PR   14-AUG-2000;   2000US-0225759.
PR   18-AUG-2000;   2000US-0226279.
PR   22-AUG-2000;   2000US-0226681.
PR   22-AUG-2000;   2000US-0226868.
PR   22-AUG-2000;   2000US-0227782.
PR   23-AUG-2000;   2000US-0227009.
PR   30-AUG-2000;   2000US-0228924.
PR   01-SEP-2000;   2000US-0229287.
PR   01-SEP-2000;   2000US-0229343.
PR   01-SEP-2000;   2000US-0229344.
PR   01-SEP-2000;   2000US-0229345.
PR   05-SEP-2000;   2000US-0229509.
PR   05-SEP-2000;   2000US-0229513.
PR   06-SEP-2000;   2000US-0230437.
PR   06-SEP-2000;   2000US-0230438.
PR   06-SEP-2000;   2000US-0231242.
PR   08-SEP-2000;   2000US-0231243.
PR   08-SEP-2000;   2000US-0231244.
PR   08-SEP-2000;   2000US-0231413.
PR   08-SEP-2000;   2000US-0231414.
PR   08-SEP-2000;   2000US-0232080.
PR   08-SEP-2000;   2000US-0232081.
PR   12-SEP-2000;   2000US-0231968.
PR   14-SEP-2000;   2000US-0232397.
PR   14-SEP-2000;   2000US-0232398.
PR   14-SEP-2000;   2000US-0232399.
PR   14-SEP-2000;   2000US-0232400.
PR   14-SEP-2000;   2000US-0232401.
PR   14-SEP-2000;   2000US-0233063.
PR   14-SEP-2000;   2000US-0233064.
PR   14-SEP-2000;   2000US-0233065.
PR   21-SEP-2000;   2000US-0234223.
PR   21-SEP-2000;   2000US-0234274.
PR   25-SEP-2000;   2000US-0234997.
PR   25-SEP-2000;   2000US-0234998.
PR   26-SEP-2000;   2000US-0235484.
PR   27-SEP-2000;   2000US-0235834.
PR   27-SEP-2000;   2000US-0235836.
PR   29-SEP-2000;   2000US-0236327.
PR   29-SEP-2000;   2000US-0236367.
PR   29-SEP-2000;   2000US-0236368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX
DR   WPI; 2001-483426/52.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Disclosure; SEQ ID NO 38125; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic antigen genomic
CC   to AAK87694 represent human immune/haematopoietic-derived cells. AAK64703
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 933 BP; 146 A; 309 C; 325 G; 153 T; 0 other;

Alignment Scores:
Pred. No.:                              1.36e-98          Length:          933
Score:                                  1243.00           Matches:         245
Percent Similarity:                     100.00%           Conservative:      0
Best Local Similarity:                  100.00%           Mismatches:        0
```

```
Query Match:    38.63%      Indels:    0
DB:             22          Gaps:      0

US-10-029-347-2 (1-625) x AAK83313 (1-933)

Qy      1   MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla   20
            |||   || ||   ||| || |  |    |  || |    | |||  ||  ||  |
Db    197   ATGCTGGCCCAGCCCCAGCGGCTGCTTCTTCATCCTGGACGGCGCGGACGAGCTGCCGGCG   256

Qy     21   LeuGlyGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg   40
            | ||| ||||  ||  ||| ||    || ||    ||  |||  ||  | |  || |  |
Db    257   CTGGGGGGGCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGGAGCGGCGCGCCGG   316

Qy     41   ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr   60
            ||   || || ||    ||   |  ||    ||  ||  || ||  || || ||  ||
Db    317   GTGCTAGGCGGGGCTGTTGCTGAGCAAGGCGCTCTGCCCACGGCCCCTGCTGGTGACCACG   376

Qy     61   ArgAlaAlaAlaProGlyArgLeuGlnGlyArgCysLysSerProGlyCysAlaGluVal   80
            || ||  || || |||  || ||   || || ||    ||||| || || || || ||
Db    377   CGGGCCGCCGCCCCCGGGAGGCTGCAGGGCCGCTGCAAGAGCCCAGGCGCGGCCGAGGTG   436

Qy     81   ArgGlyPheSerAspLysAspLysLysTyrLysTyrPheTyrLysPhePheArgAspGluArg   100
            || ||   |  |||  ||   | || ||    |  ||    || || |  || |||  || ||
Db    437   CGGGGCTTCTCCGACAAGGACAAGAAGTATTTCTACAAGTTCTTCCGGGATGAGAGG   496

Qy    101   ArgAlaGluArgAlaTyrArgPheValLysAsnGluThrLeuPheAlaLeuCysPhe   120
            || || ||  || ||    || ||   || || || || || ||  ||   || ||
Db    497   AGGGCCGAGCGGGCCTACCGTTCGTGAAGAACGAGACGCTGTTCGCGCTGTGCTTC   556

Qy    121   ValProPheValCysTrpIleValCysThrValLeuArgGlnLeuGluGlyGlyArg   140
            ||  |||   ||   ||    ||   || || || || || || || |||  || ||
Db    557   GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCTGGAGGGCGGTCGG   616

Qy    141   AspLeuSerArgThrSerThrThrSerValTyrLeuPheIleThrSerVal   160
            |||   || || ||  ||  ||  ||   || ||    ||    |  ||
Db    617   GACCTGTCGCGCACGTCCAAGACCACCAGCGTCAGCTACCTGCTT TCATCACCAGCGTT   676

Qy    161   LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCys   180
            ||    |  || || ||    ||| || || ||   || |||  |  || ||
Db    677   CTGAGCTCGGCTCCGGTCGCAGGCGGGCCCCGGCTGCAGGGCGACCTGCGCAATCTGTGC   736

Qy    181   ArgLeuAlaArgGluGlyGlyValLeuGlyValArgAlaGlnProGluLeuLysGluLeuGlu   200
            || || ||  || |||  || ||    || || || || || |  || || |||  ||  ||
Db    737   CGCCTGGCCCGCGAGGGCGTCCTGGGAGTCCGGGCGCAGCCGGAGCTGAAGGAACTGGAG   796

Qy    201   GlnLeuGluLeuArgGlySerLysSerValGlnThrArgLeuPheLeuSerLysLysGluLeuPro   220
            || || || || || |  || || || |  || || || || || |  ||  || |  || ||
Db    797   CAACTGGAGCTCGTGGCTCCAAAGTCAGACGCGTTTCTCAGCAAAAAGGAGCTGCCG   856

Qy    221   GlyValLeuGluThrGluValThrTyrGlnAsnPheIleAspGlnSerProPheGluGlnLeuPheLeu   240
            || || || ||  || || ||  || ||  ||    || || ||   || |||  || || |  || ||
Db    857   GGCGTGCTGGAGACAGAGGTCACCTACCAGAAGTTCATCGACCAGAGCTTCCAGGAGCTCCTC   916

Qy    241   AlaAlaLeuSerTyr   245
            || || || || ||
Db    917   GCGGCACTGTCCTAC   931
```

PR   29-SEP-2000;   2000US-0235367.
PR   29-SEP-2000;   2000US-0235368.
PR   29-SEP-2000;   2000US-0236369.
PR   29-SEP-2000;   2000US-0236370.
PR   02-OCT-2000;   2000US-0236802.
PR   02-OCT-2000;   2000US-0237037.
PR   02-OCT-2000;   2000US-0237038.
PR   02-OCT-2000;   2000US-0237039.
PR   02-OCT-2000;   2000US-0237040.
PR   13-OCT-2000;   2000US-0239935.
PR   13-OCT-2000;   2000US-0239937.
PR   20-OCT-2000;   2000US-0240960.
PR   20-OCT-2000;   2000US-0241221.
PR   20-OCT-2000;   2000US-0241785.
PR   20-OCT-2000;   2000US-0241786.
PR   20-OCT-2000;   2000US-0241787.
PR   20-OCT-2000;   2000US-0241808.
PR   20-OCT-2000;   2000US-0241809.
PR   20-OCT-2000;   2000US-0241826.
PR   01-NOV-2000;   2000US-0244617.
PR   08-NOV-2000;   2000US-0246474.
PR   08-NOV-2000;   2000US-0246475.
PR   08-NOV-2000;   2000US-0246476.
PR   08-NOV-2000;   2000US-0246477.
PR   08-NOV-2000;   2000US-0246478.
PR   08-NOV-2000;   2000US-0246523.
PR   08-NOV-2000;   2000US-0246524.
PR   08-NOV-2000;   2000US-0246525.
PR   08-NOV-2000;   2000US-0246526.
PR   08-NOV-2000;   2000US-0246527.
PR   08-NOV-2000;   2000US-0246528.
PR   08-NOV-2000;   2000US-0246532.
PR   08-NOV-2000;   2000US-0246609.
PR   08-NOV-2000;   2000US-0246610.
PR   08-NOV-2000;   2000US-0246611.
PR   08-NOV-2000;   2000US-0246613.
PR   17-NOV-2000;   2000US-0249207.
PR   17-NOV-2000;   2000US-0249208.
PR   17-NOV-2000;   2000US-0249209.
PR   17-NOV-2000;   2000US-0249210.
PR   17-NOV-2000;   2000US-0249211.
PR   17-NOV-2000;   2000US-0249212.
PR   17-NOV-2000;   2000US-0249213.
PR   17-NOV-2000;   2000US-0249214.
PR   17-NOV-2000;   2000US-0249215.
PR   17-NOV-2000;   2000US-0249216.
PR   17-NOV-2000;   2000US-0249217.
PR   17-NOV-2000;   2000US-0249218.
PR   17-NOV-2000;   2000US-0249244.
PR   17-NOV-2000;   2000US-0249245.
PR   17-NOV-2000;   2000US-0249264.
PR   17-NOV-2000;   2000US-0249265.
PR   17-NOV-2000;   2000US-0249297.
PR   17-NOV-2000;   2000US-0249299.
PR   17-NOV-2000;   2000US-0249300.
PR   01-DEC-2000;   2000US-0250160.
PR   01-DEC-2000;   2000US-0250391.
PR   05-DEC-2000;   2000US-0251030.
PR   05-DEC-2000;   2000US-0251988.
PR   05-DEC-2000;   2000US-0256719.
PR   06-DEC-2000;   2000US-0251479.
PR   08-DEC-2000;   2000US-0251856.
PR   08-DEC-2000;   2000US-0251868.
PR   08-DEC-2000;   2000US-0251869.
PR   08-DEC-2000;   2000US-0251989.
PR   08-DEC-2000;   2000US-0251990.
PR   11-DEC-2000;   2000US-0254097.
PR   05-JAN-2001;   2001US-0259678.

XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Rosen CA,  Barash SC,  Ruben SM;
XX

DR   WPI; 2001-483426/52.
DR   P-PSDB; AAM89960.
XX
PT   Nucleic acids encoding human immune/hematopoietic antigen polypeptides,
PT   useful for preventing, diagnosing and/or treating cancers and
PT   metastasis -
XX
PS   Claim 1; SEQ ID NO 7801; 3071pp + Sequence Listing; English.
XX
CC   AAK54951 to AAK64702 encode the human immune/haematopoietic antigen (I)
CC   amino acid sequences given in AAM82170 to AAM91921. (I) have cytostatic
CC   activity, and can be used in gene therapy and vaccine production. (I)
CC   proteins and polynucleotides may be used in the prevention, diagnosis and
CC   treatment of diseases associated with inappropriate (I) expression. For
CC   example, they may be used to treat disorders associated with decreased
CC   expression by rectifying mutations or deletions in a patient's genome
CC   that affect the activity of (I) by expressing inactive proteins or to
CC   supplement the patients own production of (I). Additionally, (I)
CC   polynucleotides may be used to produce the secreted (I), by inserting
CC   the nucleic acids into a host cell and culturing the cell to express the
CC   protein. (I) proteins and polynucleotides may be used to prevent,
CC   diagnose and treat immune/haematopoietic-related diseases, especially
CC   cancers and cancer metastases of haematopoietic-derived cells. AAK64703
CC   to AAK87694 represent human immune/haematopoietic antigen genomic
CC   sequences from the present invention. AAK54942 to AAK54950 and AAM82169
CC   represent sequences used in the exemplification of the present invention.
XX
SQ   Sequence 989 BP; 157 A; 317 C; 344 G; 164 T; 7 other;

Alignment Scores:
Pred. No.:                                                          989
Score:                      5.43e-92             Matches:           247
Percent Similarity:         1167.50              Conservative:        2
Best Local Similarity:      93.96%               Mismatches:          9
Query Match:                93.21%               Indels:              7
DB:                         .36.28%              Gaps:                1
                            22

US-10-029-347-2 (1-625) x AAK62741 (1-989)

Qy    1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla   20
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  197  ATGCTGGCCCAGCCGCAGCGGCTGCTCTTCATCCTGGACGGCGCCGACGAGCTCCCGGCG  256

Qy   21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg   40
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  257  CTGGGGGGCCCCGAGGCCGCGCCCTGCACAGACCCCTTCGAGGCGGCCAGCGGCGCGGCG  316

Qy   41  ValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThrThr   60
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  317  GTGCTAGGCGGGCTGCTCAGTAAGGCCGTGCTGCCCACGGCCCTCCTGCTGGTGACCACG  376

Qy   61  ArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGluVal   80
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  377  CGGGCCGCGGCCCCCGGGAGGCTGCAGGGCCGCCTGTGTTCCCCAGTGCGCCGAGGTG  436

Qy   81  ArgGlyPheSerAspLysAspLysLysTyrLysPhePheArgArgAspGluArg  100
         ||||||||||||||||||||||| |||| || ||:::|||||||||||
Db  437  CGCGGCTTCTCCGACAAGGACAAGAAGAGTATTTCTACAAGTATTTCCGGGATGAGAGG  496

Qy  101  ArgAlaGluArgAlaTyrArgPheValCysGluAsnGluThrLeuAsnGluThrLeuCysPhe  120
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  497  AGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACAGACCGCTGTTCGCGGCTGTGCTTC  556

Qy  121  ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGlyLeuGlyArg  140
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  557  GTGCCCTTCGTGTGTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGCTCGGTCGG  616

Qy  141  AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal  160
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db  617  GACCTGTCGCGCACGTCCAAGACCACCAGCGTGTACCTGCTTTTCATCACCAGCGTT  676

Qy  161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCys  180
         ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||

RESULT 12
ABL90582

```
ID   ABL90582 standard; cDNA; 990 BP.
XX
AC   ABL90582;
XX
DT   24-MAY-2002  (first entry)
XX
DE   Human polynucleotide SEQ ID NO 1144.
XX
KW   Cytostatic; immunosuppressive; nootropic; neuroprotective; antiviral;
KW   antiallergic; hepatotropic; antidiabetic; antiinflammatory; antiulcer;
KW   vulnerary; anticonvulsant; antibacterial; antifungal; antiparasitic;
KW   cardiant; gene therapy; cancer; immune disorder; cardiovascular disorder;
KW   neurological disease; infection; human; secreted protein; gene; ss.
XX
OS   Homo sapiens.
XX
FN   WO200190304-A2.
XX
PC   29-NOV-2001.
XX
PF   18-MAY-2001; 2001WO-US16450.
XX
PR   19-MAY-2000; 2000US-2055515P.
XX
PA   (HUMA-) HUMAN GENOME SCI INC.
XX
PI   Birse CE, Rosen CA;
XX
DR   WPI; 2002-122018/16.
DR   P-PSDB; ABB90173.
XX
PT   Novel 1465 isolated polypeptides, useful for diagnosis, treatment and
PT   prevention of neural, immune system, muscular, reproductive,
PT   gastrointestinal, pulmonary, cardiovascular, renal and proliferative
PT   disorders -
XX
PS   Claim 4; SEQ ID NO 1144; 2061pp + Sequence Listing; English.
XX
CC   The invention relates to novel genes (ABL89449-ABL90853) and proteins
CC   (ABB89040-ABB90444) useful for preventing, treating or ameliorating
CC   medical conditions e.g. by protein or gene therapy. The genes are
CC   isolated from a range of human tissues disclosed in the specification.
CC   The nucleic acids, proteins, antibodies and (ant)agonists are useful
CC   in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
CC   and ovarian cancer and other cancers of the adrenal gland, bone, bone
CC   marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
CC   (b) immune disorders e.g. Addison's disease, allergies, autoimmune
CC   haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
CC   disease, multiple sclerosis, rheumatoid arthritis and ulcerative
CC   colitis; (c) cardiovascular disorders such as myocardial ischaemias;
CC   (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
CC   epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
CC   and parasitic infections.
CC   Note: The sequence data for this patent did not form part of the
CC   printed specification, but was obtained in electronic format directly
CC   from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ   Sequence 990 BP; 158 A; 318 C; 344 G; 163 T; 7 other;
```

```
Alignment Scores:
Pred. No.:
Score:                      2.98e-90            Length:          990
Percent Similarity:         93.23%              Matches:         246
Best Local Similarity:      92.48%              Conservative:    2
Query Match:                35.66%              Mismatches:      10
DB:                         24                  Indels:          8
                                                Gaps:            1
```

```
US-10-029-347-2 (1-625) x ABL90582 (1-990)

Qy     1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   197  ATGCTGGCCCAGCCGCAGCGGCTGCTTCATCCTGGACGGCGCGGACGAGCTGCCGGCG  256

Qy    21  LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSerGlyAlaArg  40
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   257  CTGGGGGGCCCGGAGGCGGCGCCCTGCACAGACCCCTTCGAGGCGGCGAGCGGCGCGCGG  316

Qy    41  ValLeuGlyGlyLeuLeuSerLysAlaAlaLeuLeuProThrAlaLeuLeuValThrThr  60
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   317  GTGCTAGGCGGGCTGCTGAGTAAGGCGTGCTCTGCCCACGGCCCCTCCTGTGGTGACCACG  376

Qy    61  ArgAlaAlaAlaProGlyLeuGlnGlyArgLeuGlyCysSerProGlnCysAlaGluVal  80
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   377  CGGCGCCGCGCCCCCGGAGGCTGCAGGGGCCGCCTGTCCCCAGTGCGCCGAGGTG  436

Qy    81  ArgGlyPheSerAspLysLysTyrPheTyrTyrPheArgAspGluArg  100
          |||||||||||||||||||||||||||||||||||||||||||||||
Db   437  CGGGGCTTCTCCGACAAGGACAAGAAGAAGTATTTCTACAAGTWTTCCGGATCGAGAGG  496

Qy   101  ArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCysPhe 120
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   497  AGGGCCGAGCGCGCCTACCGCTTCGTGAAGGAGAACGAGACGCTGTTCGCGCTGTGCTTC  556

Qy   121  ValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGluGluGlyArg 140
          |||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   557  GTGCCCTTCGTGTGCTGGATCGTGTGCACCGTGCTGCGCCAGCAGCTGGAGCTCGGTCGG  616

Qy   141  AspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSerVal  160
          ||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db   617  GACCTGTCGCGGACCAGTCCAAGACCACCAGCGTGTACCTGCTTTTCATCACCAGCGTT  676

Qy   161  LeuSerSerAlaProValAlaAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCys 180
          |||||||||||||||||||||||||||||||||||||||||||
Db   677  CTGAGCTCGGCTCCAGTGGCGTCCGGTAGCCCGGTTGCAGGGGACCTGCGCAATCTGTGC  736

Qy   181  ArgLeuAlaArgGluGlyValLeuGlyArgArgAlaGlnPheAlaGluLeuGluGlu  200
          ||||||||||||||||||||||||||||||||||||||||||
Db   737  CGCCTGGCCCGCGAGGGCGTCCTCGGACGCCGAGCGGCGCAATTGCCGAAAAGGAACTGGAG  796

Qy   201  GlnLeuGluArgGlySer-LysValGlnThr-LeuPheLeuSerLysLysGlyLeu-  219
          ||||||||||||||| |||||| |||||  |||||||||| ||||||
Db   797  CAACTGGAGCTTCGTGGCTCCAAAAGTGCAGACCGCTGTTTCTGCAGCAAAAAGGAGCTGC  856

Qy   220  rGGlyValLeuGluThrGlyValThrTyrGlnPheIleAspGln-SerPhe-GlnGluPh  239
          |||||||| |||| |||||| |||||||  ||||| ||||| |||| |||| |||
Db   857  CCGGGGGTGCTGGAAACAGAGGCGTCACTACCAGTTCATCGACCAGAGCTTNCAGGAGTT  915

Qy   239  eLeuAlaAla--LeuSerTyrLeuLeu---GluAspGlyGlyValProArgThrAlaAl  257
          |||||||||   ||||||||||||     |||||||||||||||||  |||
Db   917  TCCTCGCGGGCCACTTGTCCTACCTTGCTGAAGGACGGCGGGGTGGCCCAAGGAACCGCC  976

Qy   258  aGly  258
          |||
Db   977  GGN  979
```

```
Db  823  TCTTTCCAGATCCTCAGGGACCCTGGTGCCTCTGCTGGGAGGAGAAACGGCCCACGGAG  882
              ||||
Qy   40  ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValThr   59

Db  883  CTGCTTCTTAACAGCTTATTCGGAAGAGCTCTCCCTGAGCTATCTTGC:CATCACC   942
Qy   60  ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu   79
              ||||    :::::
Db  943  ACACGGCCACGGCTTTGGAGAAGCTCCACCGTCTGCTGGAGCACCCCAGGCATGTGGAG  1002

Qy   80  ValArgGlyPheSerAspLysLysLysTyrLysPhePheArgAspGlu   99
              ||||||||  :::::    :::
Db 1003  ATCCTGGCCTTCTCTGAGCAGAAAGGAGAATACTTCACAAGTATTCCACAATGCA  1062

Qy  100  ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuAspPheAlaLeuCys  119
              ::::      ||||||||::
Db 1063  GAGCAGGCGGGCCAAGTCTTCAATACGTGAGGACAACGAGCCTCTTCACCATGTGC  1122

Qy  120  PheValProPheValCysTrpIleValCysThrValLeuArgGlnGlnLeuGlnLeuGly  139
              ::::::||||||||::
Db 1123  TTCGTCCCCGTGGTGTGCTGGATTGTGTGTACCGTCCTCAGCAGCAGTGGAGGGTGGG  1182

Qy  140  ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuValPheIleThrSer  159
              ||||||
Db 1183  GGGCTGTTGAGACAGAGTCCAGGACCACCAGTGTACATGCTCTACCTGCTGAGT  1242

Qy  160  ValLeuSerSerAlaProValAlaAspGlyProArgLeuGln------GlyAspLeuArg  177
              :::::
Db 1243  CTGATGAACCCCAAGCCCGGGGGGC------CCGGCCCTCCAGCCCCACCAACCAGAGA  1296

Db 1297  GGGTGTGCTCCTTGGCGGCCAGATGGGCTCTGGAATCACGAAAATCCTATTTGAGGAGCAG  1356
              ::::::::
Qy  178  AsnLeuCysArgGluAlaArgGluValGlyValLeuArgArgAlaGlnPheAlaGluLeuLys  197

Qy  198  GluLeuGluGlnLeuLeuGlyLeuGlyGlySerLysSerValGlnThrLeuProPheLeuSerLysLys  217
              :::
Db 1357  GACCTCCGGAAGCCACGACGGCCTAGACGGGGAAGACGTC--TCTGCCTCCTCAACATGAAC  1413

Qy  218  GluLeuProGlyValValLeuGluThrGlyValThrTyrGlnPheLeuAspGlnSerPheGln  237
              :::
Db 1414  ATCTTCCAGAGGACATCAACTGGAGGTACTACAGCTTCATCCACTTGAGTTCCAG  1473

Qy  238  GluPheLeuAlaAlaLeuSerTyrLeuLeuGluAspGlyGlyValProArgThrAlaAla  257
              :::::::
Db 1474  GAATTCTTTGCAGCTATGTACTATATCCTGGACGAGGGG-----------------GAG  1515

Qy  258  GlyGlyValGly------------------ThrLeuLeuArgGlyAspAlaGlnPro  270
              :::::
Db 1516  GGGGGGCAGGCCCAGGACAGTGACCAGGCTGTTGACCGGAGTACGGCGTTTCTGAA  1575

Qy  271  HisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMet  290
              :::::||
Db 1576  AGGAGCTCCTGGCACTCCACCAGCCGCTTCCTCTGTTTGGACTCCTGAACAGGACCAGG  1635

Qy  291  ArgAspIleGluArgHisPheGlyCysMetValSerGlyValAlaLysGlnGluAlaLeu  310
              :::::
Db 1636  AGCCACCTGGAGAGAGAGTCTCTGCTGGAAGGTCCGCCGCCACATCAAGATGGACCTGTTG  1695

Qy  311  ArgTrpValGlnGlyGlnGlyGlyCysProGlyCysProGlyValAlaProGluValThrGluGly  330
              ||||
Db 1696  CAGTGGATCCAAAGCAAAGCTCAG---------------------------------  1719

Qy  331  AlaLysGlyLeuAspAspThrGluProGluLeuGluLeuGluGluGlyGluProAsn  350
              :::::
Db 1720  --------------------AGCACGGCCTCCACCCTGCCAGCAGGGGTCC-------  1749

Qy  351  TyrProLeuGluLeuTyrCysLeuTyrGluThrGlnLeuGlnLeuAspAlaPheValArgGln  370
              :::
Db 1750  -----TTGGAGTTCTTCAGCTCCTGCTTGTACGAGATCCAGGAGAGAGGAGTTTATCCAGCAG  1803

Qy  371  AlaLeuCysArgPheProGluLeuAlaLeuLeuArgValValArgAlaValPheCysArgMetAspVal  390
              |||||
Db 1804  GCCCTGAGCCCGCCACTTCCAGGTGACATCGTGGTCAGCAACATTGCC---TCCAAGATGGAGCAC  1860
```

Alignment Scores:
Pred. No.:
Score:                 1.32e-61
                       826.50
Percent Similarity:    48.00%        Length:         3300
Best Local Similarity: 33.00%        Matches:         231
Query Match:           25.68%        Conservative:    105
DB:                                  Mismatches:      247
                                     Indels:          119
                                     Gaps:             17

US-10-029-347-2 (1-625) x AAL47129 (1-3300)

```
Qy    1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGlyLeu---Pro  19
              :::::
Db  763  CTCATCCGAGTTCCCAGCCGCCTCCTTTTCATCATCGACGGCTTCGATGAGCTCAAGGCC  822

Qy   20  AlaLeuGluGlyGlyProGluAlaAlaProAlaAlaProCysThrAspProPheAspAlaAlaSerGlyAla  39
              |||
Db  804  GCCCTGAGCCCGCCACTTCCAGGTGACATCGTGGTCAGCAACATTGCC---TCCAAGATGGAGCAC  1860
```

```
KW   immune system disorder; Crohn's disease; multiple sclerosis; cancer;
KW   leukaemia; autoimmune disorder; arthritis; neurological disease;
KW   Alzheimer's disease; Parkinson's disease; chromosomal mapping;
KW   tissue typing; forensic biology; predictive medicine; pharmacogenomics;
KW   transcription profiling; PYRIN-8.
XX
OS   Homo sapiens.
XX
FH   Key            Location/Qualifiers
FT   CDS            1..3186
FT                  /tag= a
FT                  /product= "Human PYRIN-8 protein #2"
XX
PN   WO200261049-A2.
XX
PD   08-AUG-2002.
XX
PF   31-JAN-2002; 2002WO-US02967.
XX
PR   31-JAN-2001; 2001US-265231P.
PR   10-SEP-2001; 2001US-318645P.
XX
PA   (MILL-) MILLENNIUM PHARM INC.
PA   (AMHP ) WYETH.
XX
PI   Bertin J, Wang W, Blatcher M;
XX
DR   WPI; 2002-627477/67.
DR   P-PSDB; AA015590.
XX
PT   New PYRIN polypeptides and nucleic acids useful for modulating and
PT   diagnosing stress-related, apoptotic and inflammatory responses, or for
PT   treating inflammatory and immune system disorders, cancers, or
PT   neurological diseases .
XX
PS   Claim 4; Fig 8; 167pp; English.
XX
CC   The invention comprises the amino acid and coding sequences of human
CC   PYRIN proteins. The PYRIN protein and DNA sequences of the invention are
CC   useful for modulating and diagnosing stress-related, apoptotic and
CC   inflammatory responses. The PYRIN protein and DNA sequences are useful
CC   for treating: inflammatory disorders and immune system disorders (e.g.
CC   Crohn's disease, reactive arthritis, multiple sclerosis, contact
CC   dermatitis, psoriasis, graft rejection, allergies, viral infections and
CC   bacterial infections); cancer (e.g. leukaemia); autoimmune disorders
CC   (e.g. systemic lupus erythematosus and arthritis); and neurological
CC   diseases (e.g. Alzheimer's disease and Parkinson's disease). The PYRIN
CC   protein and DNA sequences may also be used in screening assays, detection
CC   assays (e.g. chromosomal mapping, tissue typing or forensic biology),
CC   predictive medicine (e.g. diagnostic assays, clinical trials and
CC   pharmacogenomics) and transcription profiling. The present DNA sequence
CC   encodes a human PYRIN-8 protein.
XX
SQ   Sequence 3186 BP; 701 A; 908 C; 945 G; 632 T; 0 other;

Alignment Scores:
Pred. No.:                                   Length:          3186
Score:              1.54e-61                 Matches:          230
Percent Similarity:  48.07%                  Conservative:     106
Best Local Similarity: 32.90%                Mismatches:       247
Query Match:         25.65%                  Indels:           118
DB:                                          Gaps:              17

US-10-029-347-2 (1-625) x AAL44363 (1-3186)

Qy     1 MetLeuAlaGlnProGlnArgLeuPheLeuAspGlyAlaAspGluLeu---Pro 19
           :::::  ::::||||||||||||||||||||||||||||||||||||    ||||
Db   844 CTCATCCGAGTTCCCGAGCCGCTCCTTTCATCATCGACCGCTTCGATGAGCTCAAGCCT 903

Qy    20 AlaLeuGlyGlyProGluAlaAlaProCysThrAspProPheCysThrAspSerGlyAla 39
           |||  :::   :::::||||||||||||||||||||||||||||||||||||  |||
Db   904 TCTTTCCACGATCCTCAGGACCCTGTGCCTCTGGCGGGAGGAGAAACGGCCCCACGGAG 963
```

```
Qy      40 ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuValThr 59
           :::|||    ||   |||    |||     |||||    :|| :|||  :::|||
Db     964 CTGCTTCTTAACAGAAGTTAATTCGGAAGAGAGCTGCTCCCTGAGCTATCTTGCTCATCACC 1023

Qy      60 ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAlaGlu 79
           :::|||||    |||  ||    |||    |||     ||||||
Db    1024 ACACGGGCCCACGGCTTTGGACAAGCTCCACCGTCGCTGGAGCACCCCAGCATGTCGAG 1093

Qy      80 ValArgGlyPheSerAspLysLysLysSerTyrPheTyrLysArgPheArgAspGlu 99
           :::|||||    |||:::    |||  |||:::    |||||||
Db    1084 ATCCTGGGCTTCTCTGAGGCAGAAAGGAATACTTCTACACAGTATTCCACAATGCA 1143

Qy     100 ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys 119
           :::|||    ||    :::    |||       :::|||  ||    :::|||
Db    1144 GAGCAGGCGCGGCCAAGTCTTCAATTACTGGAGGCAACGAGCCTCTTCCACCATGTGC 1203

Qy     120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnInLeuGluLeuGly 139
           ||       :::       |||     |||:::    |||  ::: |||
Db    1204 TTCGTTCCCCCGTGGTGTGTGTGTACCTGCTCCAGCAGAGCTGGAGGTGGG 1263

Qy     140 ArgAspLeuSerArgThrSerThrThrSerValTyrLeuPheIleThrSer 159
           ::::::|||||    |||||||     |||:::    ||  ::: |||
Db    1264 GGGTGTTGAGACAGACGCCAGGACCACCACTGCAGTGTACATGCTCTACCTGCCGAGT 1323

Qy     160 ValLeuSerSerAlaProValAlaAspGlyProArgLeuGln-----GlyAspLeuArg 177
           :::||       :::    ||  |||    ||
Db    1324 CTGATGCCAACCACCGGGGGCC-----CCGCGGCCTCCAGCACCCCACCCAGAGA 1377

Qy     178 AsnLeuCysArgLeuAlaArgGluGlyValAlaArgGlyArgAlaGlnPheAlaGluLys 197
           :::||    ||    |||     |||    :::    |||||||
Db    1378 GGGTTGTGCTCCTTGGCGGCAGATGGGCTCTGGAATCAGAAAATCCTATTTGAGGAGCAG 1437

Qy     198 GluLeuGlnLeuGluLeuArgGlyGlySerLysValGlnThrLeuPheLeuSerLysLys 217
           :::||       |||    ||       :::    ||    :::
Db    1438 GACCTCCGGAAGCACGGCCTAGGACGGGAAGACGTC---TCTGCCTCCCAACATGAAC 1494

Qy     218 GluLeuProGlyValLeuGluThrGluValThrGlnSerGlnGlnSerGlnSerPheGln 237
           :::||      ||    :::    |||         :::    |||||||
Db    1495 ATCTTCCAGAAGGACAATCAACTGTAGAGGTACTACAGCTTCATCCACTTGAGTTTCCAG 1554

Qy     238 GluPheHisAlaAlaLeuSerTyrLeuLeuGluAspTyrLeuGlyGlyValProArgThrAlaAla 257
           :::|||    |||     ||       :::    |||  ||||||
Db    1555 GAA-TTCTTGCATGTACTATATATCCTGGACGAGGGG------------------GAG 1596

Qy     258 GlyGlyValGly----------ThrLeuLeuArgGlyAspAlaGlnPro 270
           :::|||
Db    1597 GGCGGGGGCAGGCCCACAGACCGGTGACCGAGCACGGCGTTTCTGAA 1656

Qy     271 HisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerArgAlaGluArgMetAla 290
           :::|||    ||       |||     |||||||
Db    1657 AGGAGCTTCCTGGACCTGCTTGTACGCAGATCCCAGACGAGGAGGAGACCCAGG 1716

Qy     291 ArgAspIleGluArgHisPheGlyCysGlyMetSerGluArgValValGlnAlaLeu 310
           :::|||||    ||       |||||||
Db    1717 AGCCACCTGGAGAAGAGTCTCGCTGGAGGTCTGCTGGGCATCCGCCGGCTGGTTG 1776

Qy     311 ArgTrpValGlnGlnGlyGlyCysProGlyValAlaProGluValThrGlyGly 330
           :::|||
Db    1777 CAGTGGATCCAAAGCAAAGCTCAG---------------------------- 1800

Qy     331 AlaLysGlyLeuAspThrGluProGluLeuGluGluGluGluProAsn 350
           :::
Db    1801 -------------------AGCGACGGCTCCACCCTGCACAGGGCTCC--- 1830

Qy     351 TyrProLeuGluLeuGluTyrCysLeuTyrGluThrGlnLeuThrGlnGluAspAlaPheValArgGln 370
           :::
Db    1831 -------TTGGAGTTCTTCAGCTGCTTGTACGCAGATCCACGAGATCCCAGGAGAGTTTATCCACCAG 1884

Qy     371 AlaLeuCysArgPheGluLeuAlaLeuGlnArgValAlaArgPheCysArgMetAspVal 390
           ::: |||
Db    1885 GCCCTGAGCCACTTCCAGGTATCCGTGTCAGCAACATTGCC--TCCAAGATGGAGCAC 1941

Qy     391 AlaValLeuSerTyrCysValArgCysValArgCysProAlaGlyAlaGlyGlnAlaLeuArgLeuIle--- 409
```

```
Db    1942 ATGGTTCCTCCGTTCTTGTCTGAAGCGCTGAGGAGGCCCAGGTGCTGCA-CTTGTATGG 2000

Qy     410 ------------------------SerCysCysArgLeu 413
Db    2001 CGCCACCTACAGCGCAGGGGAGCACCGGCGAGGCCGAGTGCTCCGAGGAGGCCACGCT 2060

Qy     414 ValAlaAlaGlnGluGlyLysLysLysSerLeuGlyLysArgLeuGlnAlaSerLeuGly 433
Db    2061 GTTGGTGCAGCTACCAGCAGGACCGTTCTGCTGGA---CGCCTACAGTGAACATCTGGC 2117

Qy     434 GlyGly---------SerSerGlnGlyThrThrLysProLeuProAla-SerLeu-- 448
Db    2118 AGCGGGCCCTGTGCACCAATCCAAACCTGATAGAGTGTCTCTGTACCGAAATGCCTGGG 2177

Qy     449 -------------LeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLeuSerSerSer 465
Db    2178 CAGCCGGGGGTGGAAGCTGCTCGTCAAGGACTCAGACACCCAACTTCAGAA 2237

Qy     465 rLeuThrLeuSerHisCysLysLysSerLeuProAspAlaValCysAspLeuSerGluAlaLe 485
Db    2238 CCTGAGGCTGAAGAGGTGCCCGATCTCCAGCTGCGAGGACCTCTCTGCAGTCT 2297

Qy     485 uArgAlaAlaProAlaAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGluAlaGl 505
Db    2298 CATAGCCAATAAGAATTGACAAGGATGGATCTCAGTGGCAACGGCGTTGGATTCCCAGG 2357

Qy     505 yLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnInCysGlyLeuGlnThrLeu 525
Db    2358 CATGATGCTGCTTTGGAGGGCCTGGCATCCCCAATGCAGTCAGATGATTCAGTT 2417

Qy     525 lGlnLeuProAspProGlnArgGly--LeuGlnTyrLeuValGlyMetLeuArgGlnSe 544
Db    2418 GAGGAAGTGTCAGCTGGATCCGGAGGCTTCGGGGATGGCTTCTGTGCTTGGCACCAA 2477

Qy     544 rProAlaLeuThrArgLeuSerLeuSerGlyCysGlnLeuProAlaProMetValThrTy 564
Db    2478 CCCACACTGGTTGAGTTGGACCTGACTGACAGGAAAATGCACTGGAGGATTTGGGCCTGAGGTT 2537

Qy     564 rLeuCysAlaValLeuGlnHisGlnGlyLeuGlyCysGlyLeuGlnThrLeu 579
Db    2538 ACTATGCCAGGGACTGAGGCACCCAGTCCAGACTGGAAGACTTTGTGGCGAAGATCTG 2597

Qy     580 SerLeuAlaSerValGluLeuSerGlyLeuSerLe 591
Db    2598 CCGCCTCACTCGCTGCCTGCCTCCAACTCCAGTGTGAACCAGAGCC 2657

Qy     591 uGlnGluLeuGlnAlaLeuLysProAspLeuValIle 606
Db    2658 GAGAGAGCTGACCTGAGCCTGAATGAGCTGGGGACCTCGGGGTGCTGCTGTGA 2717

Qy     607 ---ThrHisPro 609
Db    2718 GGGCCTCCAGGGATCCCCACGTGCAAGCTGCAAGAGCCCTGGCGGTTGGGCATCTGCCGGCTGGG 2777

Qy     610 AlaLeuAspGlyHisProGlnProPro 618
Db    2778 CTCTGCGCCCTGTGAGGGGTCTTCTGTGGGTGTCCTCCA-GGCCAACCACCAACCTCCG 2831
```

KW　glomerulonephritis; neurological disorder; Alzheimer's disease; ALS;
KW　amyotrophic lateral sclerosis; Parkinson's disease; Huntington's disease;
KW　aplastic anaemia; ischaemia; meningitis; liver disease; Crohn's disease;
KW　insulin-dependent diabetes; multiple sclerosis; Grave's disease; HIV;
KW　human immunodeficiency virus; tuberculosis; lepromatous leprosy; ss.
OS　Homo sapiens.
XX
FH　Key　　　　Location/Qualifiers
FT　CDS　　　139..3243
FT　　　　　　　/*tag= a
FT　　　　　　　/note= "Human PYRIN-1 protein"
XX
PN　WO200161005-A2.
XX
PD　23-AUG-2001.
XX
PF　16-FEB-2001; 2001WO-US040145.
XX
PR　17-FEB-2000; 2000US-0506067.
PR　01-SEP-2000; 2000US-0653901.
XX
PA　(MILL-) MILLENNIUM PHARM INC.
XX
PI　Bertin J;
XX
DR　WPI; 2001-514773/56.
DR　P-PSDB; AAEC7514.
XX
PT　Novel polypeptide comprising NBS-1 or PYRIN-1 of the pyrin domain
PT　protein family useful in screening and detection assays and for
PT　treating, e.g., cancer, viral infections, autoimmune disease, and
PT　Alzheimer's -
XX
PS　Claim 1; Fig 4; 111pp; English.
XX
CC　The invention relates to human NBS-1 nucleotide binding site) and
CC　PYRIN-1 protein and nucleic acids. NBS-1 and PYRIN-1 DNA, protein and
CC　their modulators are useful in the treatment of apoptotic and
CC　inflammatory disorders, cancer (leukaemia, melanoma, carcinoma), viral
CC　infections (including herpesvirus and adenovirus), autoimmune diseases
CC　(systemic lupus erythematosus (SLE), immune-mediated glomerulonephritis,
CC　arthritis; neurological disorders (Alzheimer's disease, Parkinson's
CC　disease, amyotrophic lateral sclerosis (ALS), retinitis pigmentosa,
CC　Huntington's disease; aplastic anaemia, ischaemia, meningitis, liver
CC　diseases. NBS-1 and PYRIN-1 DNA, protein and their modulators are also
CC　used for the treatment of inflammatory and immune disorders such as
CC　chronic inflammatory diseases such as Crohn's disease, insulin-dependent
CC　diabetes, organ specific autoimmunity, including multiple sclerosis,
CC　Grave's disease, allergy, asthma, HIV, tuberculosis and lepromatous
CC　leprosy. The present sequence is a cDNA encoding human PYRIN-1 protein.
XX
SQ　Sequence 3857 BP; 959 A; 931 C; 1023 G; 945 T; 0 other;

Alignment Scores:
Pred. No.:　　　　　　3e-59　　　　Length:　　　　　　3857
Score:　　　　　　　　800.50　　　Matches:　　　　　219
Percent Similarity:　43.64%　　　Conservative:　117
Best Local Similarity:　28.44%　　Mismatches:　　241
Query Match:　　　　24.88%　　　Indels:　　　　　193
DB:　　　　　　　　　　22　　　　　　Gaps:　　　　　　17

US-10-029-347-2 (1-625) x AAD14323 (1-3857)

Qy　　1 MetLeuAlaGlnProGlnArgLeuLeuPheLeuLeuAspGlyAlaAspGluGluProAla 20
　　　　　:::|::|:|||||||::||:|:|:||::|||||:|||:|:|::||||||||
Db　1000 ATCGTGGAGAAACCCTCCAGATCCTCTCCCATGGACGGCTCGATGAGCTGCAAGGT 1059

Qy　　21 LeuGlyGlyProGluAlaAlaPro---CysThrAspProPheGluAlaAlaSerGlyYAla 39
　　　　　:::|:|||||||:||:|||::||||||
Db　1060 GCCTTTGACGAGCACATAGGACCGCCGGCTTGCACTGACTGGCAGAAGGCCGAGCGGGGAGAC 1119

Qy　　40 ArgValLeuGlyGlyLeuLeuSerLysAlaLeuAlaLeuProThrAlaLeuLeuLeuVa-Thr 59
　　　　　:::|::||||::||

Db　1120 ATTCTCCTGAGCAGCCTCATCAGAAGAAAGCTGCTTCCGAGGCGCTTCCTGCTCATCACC 1179
　　　　　:::|||　　　　　　　　　|||　　　||||||||
Qy　　60 ThrArgAlaAlaAlaProGlyArgLeuArgLeuCysSerSerProGlnCysAlaGlu 79
　　　　　||||||
Db　1180 ACGAGACCTGTGCCCGGAGAACTGCACTTGCTGGACATCCTCGGCATGTGAG 1239

Qy　　80 ValArgGlyPheSerAspLysAspLysLysTyrTyrLysPheArgAspGlu 99
　　　　　:::|　||||||　　　　||||
Db　1240 ATCCTGGGTTCTCCGAGCCAAAGGAAGAGTACTTCTTCAGTACTTCTCCGATGAG 1299

Qy　100 ArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAlaLeuCys 119
　　　　　||||||:|
Db　1300 GCCCAAGCCAGGGCAGCCTTCAGTCAGTCTGATTCAGGAGAACAGGAGTGCCCTTCACCATGTGC 1359

Qy　120 PheValProPheValCysTrpIleValCysThrValLeuArgGlnIleLeuGlnLeuGly 139
　　　　　|||　　　|||　　||||
Db　1360 TCATCCCCCTGGTCTGCTGATCGTGTGCACTGTGCTGAGACAGCAGCAGATGGAGAGTGGC 1419

Qy　140 ArgAspLeuSerArgThrSerLysThrThrSerValTyrLeuLeuPheIleThrSer 159
　　　　　:::|||　　　||:::|||::|:||:|:|
Db　1420 AAGAGCCTTGCCCAGACATCTAAGACCACCAGCGGTGTAGGTCTTCTCCTTTCCAGT 1479

Qy　160 ValLeuSerSerAlaProValAlaAaProGlyProArgLeuGlnGlyGlyAspLeuArgAsnLeu 179
　　　　　:::|||
Db　1480 TTGCTGAGCAGCCCGGGAGGGAGCCAGGAGCACGGCCTCTGCGGCCCACCTCTGGGGGTTC 1539

Qy　180 CysArgLeuAlaArgGlyGlyValLeuGlyValArgArgAlaGlnIlePheAlaGluLysGluGluu 199
　　　　　|||:::||||||
Db　1540 TGCTCTTGGCTGCCAGATGGAATCTGGAACCAGAAAATCCTGTTTGAGGAGTCCGACCTC 1599

Qy　200 GluGlnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerLysLysGlyLleu 219
　　　　　:::|:::||
Db　1600 AGGAATCATGGACTGCAGAAGGCGGATGTG---TCTGCTTCCTGGAGGATGAACCTGTTC 1656

Qy　220 ProGlyValLeuGluThrGluValThrThrGlnPheIleAspGlnSerPheGlnGluPhe 239
　　　　　||||||
Db　1657 CAAAAGGAAGTGGACTGCGGAGAAGTTCACAGCTTCATCCACATGACTTTCCAGGAGTTC 1716

Qy　240 LeuAlaLeuSerTyrLeuLeuGluAspGlyGlyGlyValProArgThrAlaAlaGlyGly 259
　　　　　||||||　　　||||
Db　1717 TTTGCCGCCATGTACTACCTGCTGGAAGAGAAGAAAGGAAGGACGAAGGTTCCAGGG 1776

Qy　260 ---------------ValGlyThrLeuGluArgGlyAspAlaGln--- 269
Db　1777 AGTCGTTGAAGCTTCCCAGCCGAGAACGTGACAGTCCTTCTGGAAAACTATGGCAAATTC 1836

Qy　270 ProHisSerHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerArgAlaArg 289
　　　　　|||||||
Db　1837 GAAAAGGGGTATTGATTTTGTGTGACGTTTCCTCTTGGCCTGGTAAACCAGGAGAGG 1896

Qy　290 MetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluAla 309
　　　　　:::|:::|
Db　1897 ACCTCCTACTTGGAGGAAGAAGAAATTAAGTTGCAAGATCCTCAGCAAATCAGGCTGGAGCTG 1956

Qy　310 LeuArgTrpValGlnGlyGlnGlyCysProGlyValAlaProGluValThrGlu 329
　　　　　:::|:::||||
Db　1957 CTGAAATGGATT--------------------------GAAGTGAAAGCC 1980

Qy　330 GlyAlaLysAspThrGluGluProGluGluGlyGluGluGluGluGluGluPro 349
　　　　　||||||::||:::
Db　1981 AAAGCTAAAAAGCTGCAG----------ATCCAGCCCAGCCAG--- 2013

Qy　350 AsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnAspAlaPheValArg 369
　　　　　||||||　　||||||
Db　2014 ------CTGGAATTGTTCTACTGTTTGTACGAGATGATCGCAGGAGGAGGACTTCGTGCAA 2064

Qy　370 GlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArgValArgPheCysArgMetAsp 389
　　　　　:::|:::||||
Db　2065 AGGGCCATGAACTATTTCCCAAGATTGAGATCAAT-----CTCTCCACCAGAATGGAC 2118

Qy　390 ValAlaValLeuSerTyrCysValArgCysValCysProAlaGlyAlaGlyGlnAlaLeuArgLeuIle 409
　　　　　:::|:::||||

Sat Oct 4 14:15:00 2003

us-10-029-347-2.rng

Search completed: October 3, 2003, 13:58:45
Job time : 475 secs

Db  2119  CACATGGTTCTTCCTTTGCATTGAGAACTGTCATCGGTGGAGTCACTGTCCCTGGGG  2178

Qy  410  SerCysArgLeuValAlaAlaGlnGluLysLysLysSerLeuGlyLysArgLeuGln  429
              :::|:|||            :::::|||::::::            |||
Db  2179  TTTCTCCATAACATGCCCAAGGAGGAGGAGGAAAAGGAGGCCGACCACCTTGAT  2238

Qy  430  ---------------AlaSerLeuGlyGlySerSerGlnGly---ThrThr  441
                              ::|:||            |||  |||
Db  2239  ATGGTGCAGTGTGTCTCCCTAGCTCCTCATGCTGCCTGTTCTCATGGATTGGTGAAC  2298

Qy  442  LysGlnLeuProAlaSerLeuHisProLeuPhe-----------GlnAla  455
              |||            :::|::|            |||::|
Db  2299  AGCCACCTCACTTCCAGTTTTGCCGGGGCCTCTTTTCAGTTCTGAGCACCAGCCAGAGT  2358

Qy  456  MetThr------------------------------AspPro  459
              :::|:                              |||
Db  2359  CTAACTGAATTGACCTCAGTGACAATTCTCTGGGGGACCCAGGGATGAGAGTGTTGTGT  2418

Qy  459  ----------------------------------------------------  459

Db  2419  GAAACGCTCCAGCATCCTGGCTGTAACATTCGGAGATTGTGGTTGGGGCGCCTGTGGCCTC  2478

Qy  459  ----------------------------------------------------  459

Db  2479  TCCATGAGTGCTGCTTCGACATCTCCTTGGTCTTCAGCAGCAACCAGAGCTGGTGGAG  2538

Qy  459  ----------------------------------------------------  459

Db  2539  CTGGACCTGAGTGGACAACGCCCCTCGGATGACTTCGGAATCAGACTTCTGTGTGGGACTG  2558

Qy  460  ---------LeuCysHisLeuSerSerLeuThrLeuSerHisCysLysLeuProAspAla  476
                   ::::|||            |||            |||            |||
Db  2599  AAGCACCTGTTGTGCAATCTGAAGAAGCTTCGTTGGCCAGCCGCTGCCTCACATCAGCA  2658

Qy  477  ValCysArgAspLeuSerGluAlaLeuArgAlaAlaProAlaLeuThrGlu---LeuGlyLeu  495
              ::::|||:::|||            :::            ::::            :::
Db  2659  TGTTGTCAGGATCTTGCATCGTATGAGCACCAGCCATTCCCTGACCAGACCCTATGTG  2718

Qy  497  LeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeuSerGlnGlyLeuLeuAlaTrpPro  516
              :::::|||            :::|:|            |::            |::
Db  2719  GGGGAGAATGCCCTGGGAGACTCAGGAGTCGCAATTTTATGTGAAAAAGCCAAGAATCCA  2778

Qy  517  GlnCysArgValGlnThrValArgVal-----  525
              :||||            :::
Db  2779  CAGTGTAACCTGCGAGAAACTGGGGTTGGTGAATTCTGGCCTTACGTGCTCTGTTGTTCA  2838

Qy  525  ---------------------------------------------------  525

Db  2839  GCTTTGTCCTCGGTACTCAGCACTAATCAGAATCACGCACCTTTACTGCGAGGCAAC  2898

Qy  526  GlnLeuProAspProGlnArgGlyLeuGlnTyrLeuVal---GlyMetLeuArgGlnSer  544
              :||:::            ::::||            :::            :::::
Db  2899  ACTCTCGGAGAC-----AAGGGGATCAAACTACACTCTGCGAGGGACTCTTGCACCCCGAC  2952

Qy  545  ProAlaLeuThrLeuAspLeuSerGlyCysGlnLeuProAlaProMet----------  561
              :::|||            :::            |||:::
Db  2953  TGCAAGCTTCAGGTGTTGAATTAGACAACTGCAACCTTCAGTCACACTGCTGCTGGGAT  3012

Qy  561  ----------------------------------------------------  561

Db  3013  CTTTCCCACTTCTGACCTCCAGCCAGGCCTGCGAAAGCTGAGCCTGGCAACAATGAC  3072

Qy  562  ----------ValThrTyrLeuCysAlaValLeuGlnHisGlnGlyCysG lyLeu  576
                        |||            :::|||            |||::::
Db  3073  CTGGGCGACCTGGGGGTCATGATGTTCTGTGAAGTGCTGAAACAGCAGCTGCCTCCTG  3132

Qy  577  GlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGluGlnLeuGlnAla  596
              ::::|||            :::|::            ::::::            |||:
Db  3133  CAGAACCTGGGGTTGTCTGAAAATGTATTTCAATTATGAGACAAAAAGTGCGTTAGAAACA  3192

Qy  597  ValLysArgAlaLysProAspLeuValIle  606
              :::|||            :::
Db  3193  CTTCAAGAAGAAAAAGCCTGAGCTGACCGTC  3222

GenCore version 5.1.6
Copyright (c) 1993 - 2003 Compugen Ltd.

QM protein - nucleic search, using frame_plus_p2n model

Run on:     October 3, 2003, 13:50:45 ; Search time 101 Seconds
                                         (without alignments)
                                         2731.331 Million cell updates/sec

Title:          US-10-029-347-2
Perfect score:  3218
Sequence:       1 MLAQPQRLFILDGADELPA.........ITHPALDGHPQPPKELISTF 625

Scoring table:  BLOSUM62
                Xgapop 10.0 , Xgapext    0.5
                Ygapop 10.0 , Ygapext    0.5
                Fgapop  6.0 , Fgapext    7.0
                Delop   6.0 , Delext     7.0

Searched:       569978 seqs, 220690566 residues

Total number of hits satisfying chosen parameters:    139956

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
                 Maximum Match 100%
                 Listing first 45 summaries

Command line parameters:  -DEV=xlh
-MODE=frame+_p2n.model -DEV=xlh
-Q=/cgn2_1/USPTO_spool/US10029347/runat_02102003_173437_17096/app.query.fasta_1_775
-DB=Issued_Patents_NA -QFMT=fastap -SUFFIX=rni -MINMATCH=0.1 -LOOPCL=0
-LOOPEXT=0 -UNITS=bits -START=1 -END=-1 -MATRIX=blosum62 -TRANS=human40.cgi
-LIST=45 -DOCALIGN=200 -THR_SCORE=pct -THR_MAX=100 -THR_MIN=0 -ALIGN=15
-MODE=LOCAL -OUTFMT=pto -NORM=ext -HEAPSIZE=500 -MINLEN=0 -MAXLEN=2000000000
-USER=US10029347 @CGN_1_1_56 @runat_02102003_173437_17096 -NCPU=6 -ICPU=3
-NO_MMAP -LARGEQUERY -NEG_SCORES=0 -WAIT -DSPBLOCK=100 -LONGLOG
-DEV_TIMEOUT=120 -WARN_TIMEOUT=30 -THREADS=1 -XGAPOP=10 -XGAPEXT=0.5 -FGAPOP=6
-FGAPEXT=7 -YGAPOP=10 -YGAPEXT=0.5 -DELOP=6 -DELEXT=7

Database :  Issued_Patents_NA:*
            1: /cgn2_6/ptodata/2/ina/5A_COMB.seq:*
            2: /cgn2_6/ptodata/2/ina/5B_COMB.seq:*
            3: /cgn2_6/ptodata/2/ina/6A_COMB.seq:*
            4: /cgn2_6/ptodata/2/ina/6B_COMB.seq:*
            5: /cgn2_6/ptodata/2/ina/PCTUS_COMB.seq:*
            6: /cgn2_6/ptodata/2/ina/backfiles1.seq:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | % Query Match | Length | DB | Match | ID | Description |
|---|---|---|---|---|---|---|---|
| 1 | 680.5 | 21.1 | 678 | 4 | | US-09-016-434-1C11 | Sequence 1011, Ap |
| 2 | 400 | 12.4 | 257 | 4 | | US-09-016-434-208 | Sequence 208, App |
| 3 | 325 | 10.1 | 4543 | 2 | | US-08-519-547A-5 | Sequence 5, Appli |
| 4 | 312.5 | 9.7 | 4441 | 1 | | US-09-641-999-2 | Sequence 2, Appli |
| 5 | 304.5 | 9.5 | 3393 | 1 | | US-08-295-502-1 | Sequence 1, Appli |
| 6 | 304.5 | 9.5 | 3393 | 5 | | PCT-US95-10691-1 | Sequence 1, Appli |
| 7 | 289 | 9.0 | 2859 | 4 | | US-09-099-041A-9 | Sequence 9, Appli |
| 8 | 289 | 9.0 | 2859 | 4 | | US-09-245-281-9 | Sequence 9, Appli |
| 9 | 289 | 9.0 | 2859 | 4 | | US-09-207-359B-9 | Sequence 9, Appli |
| 10 | 289 | 9.0 | 2859 | 4 | | US-09-340-620A-9 | Sequence 9, Appli |
| 11 | 289 | 9.0 | 3382 | 4 | | US-09-099-041A-7 | Sequence 7, Appli |
| 12 | 289 | 9.0 | 3382 | 4 | | US-09-245-281-7 | Sequence 7, Appli |
| 13 | 289 | 9.0 | 3382 | 4 | | US-09-207-359B-7 | Sequence 7, Appli |
| 14 | 289 | 9.0 | 3382 | 4 | | US-09-340-620A-7 | Sequence 7, Appli |
| 15 | 285 | 8.9 | 4302 | 4 | | US-09-245-281-38 | Sequence 38, Appl |
| 16 | 285 | 8.9 | 4302 | 4 | | US-09-207-359B-38 | Sequence 38, Appl |
| 17 | 285 | 8.9 | 4302 | 4 | | US-09-340-620A-38 | Sequence 38, Appl |
| 18 | 251.5 | 7.8 | 4141 | 4 | | US-09-245-281-42 | Sequence 42, Appl |
| 19 | 251.5 | 7.8 | 4141 | 4 | | US-09-207-359B-42 | Sequence 42, Appl |
| 20 | 251.5 | 7.8 | 4141 | 4 | | US-09-340-620A-42 | Sequence 42, Appl |
| 21 | 220.5 | 6.9 | 32042 | 4 | | US-09-245-281-44 | Sequence 44, Appl |
| 22 | 220.5 | 6.9 | 32042 | 4 | | US-09-340-620A-63 | Sequence 63, Appl |
| 23 | 216 | 6.7 | 1371 | 2 | | US-08-910-731-1 | Sequence 1, Appli |
| 24 | 216 | 6.7 | 1371 | 2 | | US-08-795-395-1 | Sequence 1, Appli |
| 25 | 214 | 6.7 | 1371 | 2 | | US-08-910-731-7 | Sequence 7, Appli |
| 26 | 212 | 6.6 | 1386 | 2 | | US-08-910-731-5 | Sequence 5, Appli |
| 27 | 209.5 | 6.5 | 1374 | 2 | | US-08-910-731-3 | Sequence 3, Appli |
| 28 | 209.5 | 6.5 | 1374 | 2 | | US-08-795-395-3 | Sequence 3, Appli |
| 29 | 154.5 | 4.8 | 3080 | 4 | | US-09-099-041A-25 | Sequence 25, Appl |
| 30 | 154.5 | 4.8 | 3080 | 4 | | US-09-245-281-25 | Sequence 25, Appl |
| 31 | 154.5 | 4.8 | 3080 | 4 | | US-09-207-359B-25 | Sequence 25, Appl |
| 32 | 154.5 | 4.8 | 3080 | 4 | | US-09-340-620A-25 | Sequence 25, Appl |
| 33 | 151 | 4.7 | 1470 | 4 | | US-09-099-041A-27 | Sequence 27, Appl |
| 34 | 151 | 4.7 | 1470 | 4 | | US-09-245-281-27 | Sequence 27, Appl |
| 35 | 151 | 4.7 | 1470 | 4 | | US-09-207-359B-27 | Sequence 27, Appl |
| 36 | 151 | 4.7 | 1470 | 4 | | US-09-340-620A-27 | Sequence 27, Appl |
| 37 | 137.5 | 4.3 | 3116 | 4 | | US-09-904-615-43 | Sequence 43, Appl |
| 38 | 132.5 | 4.1 | 1876 | 4 | | US-09-620-405B-477 | Sequence 477, App |
| 39 | 132.5 | 4.1 | 1876 | 4 | | US-09-604-287A-477 | Sequence 477, App |
| 40 | 123 | 3.8 | 2072 | 4 | | US-09-016-434-1181 | Sequence 1181, Ap |
| 41 | 120 | 3.7 | 2346 | 3 | | US-09-193-503B-5 | Sequence 5, Appli |
| 42 | 120 | 3.7 | 2346 | 4 | | US-09-415-839-5 | Sequence 5, Appli |
| 43 | 119 | 3.7 | 2196 | 4 | | US-09-252-991A-2198 | Sequence 2198, Ap |
| 44 | 117.5 | 3.7 | 1518 | 4 | | US-09-620-405B-484 | Sequence 484, App |
| 45 | 117.5 | 3.7 | 1518 | 4 | | US-09-604-287A-484 | Sequence 484, App |

ALIGNMENTS

RESULT 1
US-09-016-434-1C11
; Sequence 1011, Application US/09016434
  Patent No. 6500938
; GENERAL INFORMATION:
;  APPLICANT: Janice Au-Young
;  APPLICANT: Jeffrey J. Seilhamer
;  TITLE OF INVENTION: COMPOSITION FOR THE DETECTION OF SIGNALING
;  TITLE OF INVENTION: PATHWAY GENE EXPRESSION
;  NUMBER OF SEQUENCES: 1490
;  CORRESPONDENCE ADDRESS:
;   ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
;   STREET: 3174 PORTER DRIVE
;   CITY: PALO ALTO
;   STATE: CALIFORNIA
;   COUNTRY: USA
;   ZIP: 94304
;  COMPUTER READABLE FORM:
;   MEDIUM TYPE: Floppy disk
;   COMPUTER: IBM PC compatible
;   OPERATING SYSTEM: PC-DOS/MS-DOS
;   SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
;  CURRENT APPLICATION DATA:
;   APPLICATION NUMBER: US/09/016,434
;   FILING DATE: HEREWITH
;   CLASSIFICATION:
;  PRIOR APPLICATION DATA:
;   APPLICATION NUMBER:
;   FILING DATE:
;   CLASSIFICATION:
;  ATTORNEY/AGENT INFORMATION:
;   NAME: Zeller, Karen J.
;   REGISTRATION NUMBER: 37,071
;   REFERENCE/DOCKET NUMBER: PA-0002 US
;   TELECOMMUNICATION INFORMATION:
;    TELEPHONE: (650) 855-0555

TITLE OF INVENTION:   COMPOSITION FOR THE DETECTION OF SIGNALING
TITLE OF INVENTION:   PATHWAY GENE EXPRESSION
NUMBER OF SEQUENCES: 1490
CORRESPONDENCE ADDRESS:
    ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
    STREET: 3174 PORTER DRIVE
    CITY: PALO ALTO
    STATE: CALIFORNIA
    COUNTRY: USA
    ZIP: 94304
COMPUTER READABLE FORM:
    MEDIUM TYPE: Floppy disk
    COMPUTER: IBM PC compatible
    OPERATING SYSTEM: PC-DOS/MS-DOS
    SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
    APPLICATION NUMBER: US/09/016,434
    FILING DATE: HEREWITH
    CLASSIFICATION:
PRIOR APPLICATION DATA:
    APPLICATION NUMBER:
    FILING DATE:
    CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
    NAME: Zeller, Karen J.
    REGISTRATION NUMBER: 37,071
    REFERENCE/DOCKET NUMBER: PA-0002 US
TELECOMMUNICATION INFORMATION:
    TELEPHONE: (650) 855-0555
    TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 208:
SEQUENCE CHARACTERISTICS:
    LENGTH: 257 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
IMMEDIATE SOURCE:
    LIBRARY: SINTBST01
    CLONE: 1499408
US-09-016-434-208

Alignment Scores:
Pred. No.:                                    Length:        257
Score:                  2.71e-33              Matches:        81
Percent Similarity:     400.00                Conservative:    1
Best Local Similarity:  96.47%                Mismatches:      3
Query Match:            95.23%                Indels:          1
DB:                     12.43%                Gaps:            0
                        4

US-10-029-347-2 (1-625) x US-09-016-434-208 (1-257)

Qy   91  TyrPheTyrLysPhePheArgAspGluArgArgAlaGluArgAlaTyrArgPheValLys  110
             |||||||||::|||||||||:|||||||||||:|:||:|:|||||||||
Db    2  TATTTCTACAAGTTCTTCCGGGATGAGAGAGGAGGCGGAGGCGGCCTACCGCTTCGTGAAG  61

Qy  111  GluAsnGluThrLeuPheAlaLeuCysPheValProPheValCysTrpIleValCysThr  130
             |||||||||| ||:|||||| ||||||||||||||||||:||||||||||||||||
Db   62  GAGAACGAGAGCGCTGTTCGCGCTGTGCTTCGTGCCCTTCGTGTGCTGGATCGTGTGCACC  121

Qy  131  ValLeuArgGlnGlnLeuGluLeuGluArgAspLeuSerArgThrSerLysThrThrThr  150
             |:||||||||::|||||||||:|||||||||:||||::||:||||||||||||
Db  122  GTGCTGCGCCA-NANTGNAGCGTCGGTCGGGACCGTCGCGCCAGACCACCACG  180

Qy  151  SerValTyrLeuLeuPheIleThrSerValLeuSerSerAlaProValAlaAspGlyPro  170
             ||||||||||||||||||||||||||||||||||||||::|||||||
Db  181  TCAGTGTACCTGCTTTTCATCACCAGCGGTTCTGAGCTCGGCTCCGGTAGCGACGGGCCC  240

Qy  171  ArgLeuGlnGlyAsp  175
             |||||||||||
Db  241  CGGTTGCAGGGCGAC  255

RESULT 3
US-08-519-547A-5

TELEFAX: (650) 845-4166
INFORMATION FOR SEQ ID NO: 101::
SEQUENCE CHARACTERISTICS:
    LENGTH: 678 base pairs
    TYPE: nucleic acid
    STRANDEDNESS: single
    TOPOLOGY: linear
IMMEDIATE SOURCE:
    LIBRARY: BRAINCTO4
    CLONE: 927003
US-09-016-434-1011

Alignment Scores:
Pred. No.:                                    Length:        678
Score:                  5.19e-62              Matches:       177
Percent Similarity:     680.50                Conservative:    1
Best Local Similarity:  81.28%               Mismatches:      36
Query Match:            80.82%               Indels:         11
DB:                     22.5%                 Gaps:            0
                        4

US-10-029-347-2 (1-625) x US-09-016-434-1011 (1-678)

Qy  108  PheValLys-GluAsnGluThrLeuPheAlaLeuCysPheValProPheVa CysTrpI  127
             |||||||| ::|| ||:|||||||||||||||:|||||||||||||||||||
Db   36  TTCGTGAAGNGNCGCTCTANACCATANCNCNCTGTTCTTCGTGCCCTTCGTGTGCTGGAT  95

Qy  127  evalCysThrValLeuArgGlnGlnLeuGluLeuGluLeuGlyArgAspLeuSerArgThrSerLy  147
             :|||||||||||||||||||||||||||||||:|||:||||||||||||
Db   96  CGTGTGCACCGTNCTGCGGCAGCAACTGGAGCTCGGTCGGACCGTCGCGCACGTCCAA  155

Qy  147  s-ThrThrThrSerValTyrLeuLeuPheIleThrSerValLeuSerSerAlaProValA  167
             |||||||||||||||||||||||||||||||||||||||||||||||:|||||
Db  156  GGACCACCACGTCAGTGTACCTGCTTTTCATCACCAANNNTTCTGAGCTCGGCTCCGG:AG  215

Qy  167  laAspGlyProArgLeuGlnGlyAspLeuArgAsnLeuCysArgLeuAlaArgGluGlyV  187
             ||||||||||||||||||||| |||||||||||:|||||||:||:||||||||||
Db  216  CCGACGGGCCCCGGTTGCAGGGCGACCTGCGCAATCTGTGCCGCCTGGCCCGCGAGGGCG  275

Qy  187  alLeuGlyArgArgAlaGlnPheAlaGluLysGlyGluGlnLeuGluGlnLeuGlnLeuGluLeuArgGly  207
             |||||||||||||||||||:|||||||||||||||||||||||||:||||||
Db  276  TCCTCGGACGCCAGGGCGCANTTGCCGAGAAGGAACTGGAGCTTCGTGGCT  335

Qy  207  erLysValGlnThrLeuPheLeuSerLysLysGluSerGluProGlyValLeuGluThrGluV  227
             ||||||||||||||||| |||| ||||:||||||||||||:||||||||||||||
Db  336  CCAAAGTGCAGACGCTGTTTCTCANCAAAAGGA-CTGCCGGGGTGCTGGAAACAGAGG  394

Qy  227  alThrTyrGlnSerPheAsp-AspGlnSerPheGlnSerPheGluei-AlaAlaLeuSerTyrLe  246
             ||||||||||||||| |||||||||:| :||::|||:|||||||||||:
Db  395  TCACCTACCAGTTCATCGGACCCAGAGCTTCNAGGAATTCC:CCGGCGCACCGTCCTACC:  454

Qy  246  uLeuGluAspGlyGlyValProArgThrArgAlaAlaGlyGlyVal-GlyThrLeuLeuArgG  266
             |:|||||:||||||||||||||||:||||||||||:||||| |||||||||||
Db  455  GCTGGAAGACGNGGGGTGCCCA:GACCGCCGCTGCGCGGC:TTGGACACTCCTGCT::  514

Qy  266  lyAspLeuArgHisSerHisLeuValLeuThrArgPheLeuPheGlyLeuLeu  286
             :||:|||:|||| |||||||:|||||||||||||||||
Db  515  GGGACGCCCA-CCGCACAGCCACTGGTGC:CAC-ACCCGC::CCTTCCGACTGCTGA  572

Qy  286  erAlaGluArgMetArgAspIleGluGluArgHisPheGlyCysMetValSerGluArgValL  306
             ||:||||||||||:||||||||||||:|||||||||:|||||||:|||:
Db  573  ANGNTGAGCGAAG-CGNNACATCNACCCCCATTC-CGCTGCATGGTTTCCAACNTTTTA  630

Qy  306  ysGlnGluAlaLeuArgTrpValGlnGlyGlnGlyGlyCysPro  321
             :||:|||||||||:|||||||||||||
Db  631  ACCA-GAAGCCCGTGCGGTGGGTTTCAGGAACAGGANTNGCNCCCCG  676

GENERAL INFORMATION:
    APPLICANT: Janice Au-Young
    APPLICANT: Jeffrey J. Seilhamer

RESULT 2
US-09-016-434-205
Sequence 208 Application US/09016434
Patent No. 6500938

Sequence 5, Application: US/08519547A
Patent No. 5994082
GENERAL INFORMATION:
APPLICANT:
TITLE OF INVENTION:  Proteins Essential for the Expression of
TITLE OF INVENTION:  Vertebrate MHC Class II Genes, DNA Sequences Encoding Same
TITLE OF INVENTION:  and Pharmaceutical Compositions
NUMBER OF SEQUENCES: 6
CORRESPONDENCE ADDRESS:
 ADDRESSEE: FISH & NEAVE
 STREET: 1251 AVENUE OF THE AMERICAS
 CITY: NEW YORK
 STATE: NEW YORK
 COUNTRY: U.S.A.
 ZIP: 10020-1104
COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: MS-DOS
 SOFTWARE: WordPerfect 6.1
CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/519,547A
 FILING DATE: 25-AUG-1995
 CLASSIFICATION: 435
PRIOR APPLICATION DATA:
 APPLICATION NUMBER: EP94115378.7
 FILING DATE: 26-AUG-1994
ATTORNEY/AGENT INFORMATION:
 NAME: HALEY, JAMES F.
 REGISTRATION NUMBER: 27,794
 REFERENCE/DOCKET NUMBER: VOS-11
TELECOMMUNICATION INFORMATION:
 TELEPHONE: 212-596-9000
 TELEFAX: 212-596-9090
INFORMATION FOR SEQ ID NO: 5:
 SEQUENCE CHARACTERISTICS:
 LENGTH: 4543 base pairs
 TYPE: nucleic acid
 STRANDEDNESS: double
 TOPOLOGY: linear
 MOLECULE TYPE: DNA (genomic)
 ANTI-SENSE: NO
 HYPOTHETICAL: NO

US-08-519-547A-5

Alignment Scores:

| | | | |
|---|---|---|---|
| Pred. No.: | | Length: | 4543 |
| Score: | 2.41e-23 | Matches: | 195 |
| Score: | 325.00 | Conservative: | 82 |
| Percent Similarity: | 36.93% | Mismatches: | 296 |
| Best Local Similarity: | 26.00% | Indels: | 179 |
| Query Match: | 10.10% | Gaps: | 29 |
| DB: | 2 | | |

US-10-029-347-2 (1-625) x US-08-519-547A-5 (1-4543)

```
Qy    1 MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGlyAlaAspGluLeuProAla     20
          ::|||  :::|||  :::|||  :::|||  :::|||
Db 1574 ATCCTGAAGAGAACCTGACCCGTTCTGCTCATCCTAGACGCCCTCGAGGAGCTGGAAGCG 1633

Qy   21 LeuGlyGlyProGluAlaAlaProCys---ThrAspProPheGluAlaAlaSerGlyAla   39
          :::|||  :::|||  :::
Db 1634 CAAGATGGCTTCCTGCCACAGCACGTGCCGGCCACCGGCGAGCCCTGCTCCTCCGG     1693

Qy   40 ArgValLeuGlyValGlyLeuLeuSerLysAlaLeueuProThrAlaLeuLeuuValThr   59
          ::|||  :::|||  :::|||  :::
Db 1694 GGGCTGGTGGCCGGCCCTTTCCAGAAGAAGCTGCTCCAGGAGGTTGCACCCTCCTCACA  1753

Qy   60 ThrArgAlaAlaAlaProGlyArgLeuGlnLyuArgLeuCysSerProGlnCysAla---   78
          ::|||  :::|||
Db 1754 GCCCGCCCCGG-----GGCCGCCTGGTCCAGAGCCTGAGCAAGGCCGACGCCCTATTT  1867

Qy   79 GluValArgGlyPheSerAspLysLysAspLysLysLysTyrPheTyrLysPheArgAsp   98
          :||:::  ::|||  :::  ||||||:
Db 2760 TCACCCGTTTCAGGGCTGCCTTGAGCGACACAGGTGGCGCTGTGGGAGTCCCTGCGCAGC 2819
```

Db 1808 GAGCTGTCCGGCTTCCATGGAGCAGGCCCAGGCATACGTGATGGCTACTTTGAGAGC 1867
Qy   99 GluArgArgAlaGlu-------ArgAlaTyrArgPheValLysGluAsnGluThrLeu 115
Db 1868 TCAGGGATGACAGAGCACCAAGACAGAGCCCTGACGGTCCTCCGGACCGGCCACTTCTT 1927
Qy  116 PheAlaLeuCysPheValProPheValCysTrpIleValCysThrValLeuArgGlnGln 135
Db 1928 CTCAGTCACAGCCACAGCCCTACTTTGTGCCGGCAGTGTGCCAGCTCTCAGAGGCCCTG 1987
Qy  136 LeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu--    154
Db 1988 CTGGAGCTTGGGAGGAC---GCCAAGTGCCCTCCACGCTCACGGGACTCTATGTCGGC 2044
Qy  155 LeuPheIleThrSerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGly 174
Db 2045 CTGCTGGGCGTGCAGCCCTCGACAGCCCCCC-----------GGG 2080
Qy  175 AspLeuArgAsnLeuCysArgLeuAlaArgGluValValLeuGlyValAlaArgAlaGlnPhe 194
Db 2081 GCCCTGGCAGAGCTGGCCAAGCTGGCCTGGGAG-------CTGGGCCGCAGCACTCAA-- 2131
Qy  195 AlaGluLysGluLeuGluLysGlyLeuGlnLeuArgGlySerLysValGlnThrLeuPheLeu 214
Db 2132 ------AGTACCCTACAGGAGGACCAGTTCCCATCCGGCAGACGTGAGGACCTGGGCGATG 2185
Qy  215 SerLys-------LysGluLeuProGlyValLeuGluThrGluValThrTyr---Gln 230
Db 2186 GCCAAAGCTAGTCCAACACCACCGGGCGCCAGAGTCCGAGTTGGCCTTCCCCAGC 2245
Qy  231 PheIleAspGlnSerPheGlnGlnSerPheGluPheLeu--AlaAlaLeuSerTyrLeuLeuGluGluAsp 249
Db 2246 TTCCTCGTCCAATGCTTCCTGGGGGCCCTGTGCTGGCTCTGAGTGGCGAAATCAAGGAC 2305
Qy  250 GlyGlyValProArgThrAlaAlaGlyGlyValGlyTyrThrLeuLeuArgGlyValAspAlaGln 269
Db 2306 AAGGAGCTCCCCAGTACCTAGCA------TTGACCCCAAGGAGGAAGAGG 2350
Qy  270 ProHisSerHis---LeuValLeuThrThrArgPheGlyPheLeuPheGlyLeuLeuSerAlaGlu 288
Db 2351 CCCTATGACAACTGGCTGGAGGGCGGTGCCACGCCTTCTGGCTGGGCTGATCTTCCAGCCT 2410
Qy  289 ArgMetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGlu 308
Db 2411 CCCGGCGCCTGCCTGGAGCCCTACTCGGCCATGGGCTGCCTCGGTGGA-CAGGAA 2463
Qy  309 AlaLeuArgTrpValGlnGlyGlnGlyCys--ProGlyValAlaProGluValThr 328
Db 2470 GCGCAGA-AGTGGCTGCCGAGGTACCTGAAGGCGGTGCAGCCGGGG---------A 2513
Qy  328 hrGluGlyAlaLysGluLeuGluAspThrGluGlyGluGluGluGlyGlyGlyGlu 348
Db 2514 CACTGCGGGCGGCCGCAGCTG 2533
Qy  348 luProAsnTyrProLeuGluLeuCysLeuTyrCysLeuTyrGluTyrGlnGluAspAlaPheV 368
Db 2534 ------CTTGAGCTGCTACTGCGCCCACGAGGCCCGGAGGCTGGAATTT 2579
Qy  368 alArgGlnAlaAlaLeuCysArgPhePro---GluLeuAlaLeuGlnArgValArgPheCysA 387
Db 2580 GGCAGCACGTGGTACAGGAGCTCCCCGGCCGCCTCTTTTTCTGGGACCCCTCACGC 2639
Qy  387 rgMetAspValAlaValAlaLeuSerTyrCysValArgCys--- 399
Db 2640 CTCCTGATGCACATGTACTGGGACAAGGCCTTGGAGGCGGCCAAGACTTCTCCTGG 2699
Qy  400 --------CysProAlaGly---GlnAlaLeuArgLeuIleSerCys-- 411
Db 2700 ACCTCCGCAGCACTGGCCATTTGCCCCTGGATTGGGAGCCTCGTGGACTCAGCTGTG 2759
Qy  411 --- 411
Db 2760 TCACCCGTTTCAGGGCTGCCTTGAGCGACACAGGTGGCGCTGTGGGAGTCCCTGCGCAGC 2819

```
; NUMBER OF SEQ ID NOS: 6
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 2
; LENGTH: 4441
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-641-999-2

Alignment Scores:
Pred. No.:                4.76e-22        Length:        4441
Score:                    312.50          Matches:       196
Percent Similarity:       37.65%          Conservative:  86
Best Local Similarity:    26.17%          Mismatches:    289
Query Match:              9.71%           Indels:        181
                                          Gaps:          29
DB:

US-10-029-347-2 (1-625) x US-09-641-999-2 (1-444:)
```

```
Qy    1  MetLeuAlaGinProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  2C
            ::||||   ||:|| :|| |||  |||| |||||
Db 1482  ATCTTGAAGAGACCTGACCGCGTTCTGCTCATCCTAGACGCGTTCGAGGAGCTTGGAAGCG  1541

Qy   21  LeuGlyGlyProGluAlaAlaProCys---ThrAspProPheGluAlaAlaSerGlyAla  39
Db 1542  CAAGATGGCTTCCTGCACAGCACGTGCGACACGGCGGACCCCGCTCCTCCGG  1601

Qy   40  ArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeuValThr  59
Db 1602  GGGCTGCTTGGCCGGCCTTTTCACAGAAGCTGCTCCGAGGTTGCACCCTCCTCCTCACA  1661

Qy   60  ThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProLeuCysAla---  78
Db 1662  GCCCGGCCCCGG---------GGCCGCTGGTCCAGAGCCTGAGCAAGGCCGACGCCCTATT  1715

Qy   79  GluValArgPheSerAspLysAspLysLysLysTyrPheLysPheProPheArgAsp  98
Db 1716  GAGCTGTCCGGCTTCTCCATGGAGCAGGCCAGGCATACGTGATGACGTC  1775

Qy   99  GluArgArgAlaGlu---------ArgAlaTyrArgPheValLysGluAsnGluThrLeu  115
Db 1776  TCAGGGATGACAGAGCACCAAGACAGAGCCCTCCTCCGGGACCGGCCACTTCTT  1835

Qy  116  PheAlaLeuCysPheValProPheValCysTrpIleValCysThrValLeuArgValGln:  135
Db 1836  CTCAGTCACAGCACCAGAGCCTACTTTGTGCCGCCAGTGTGCCAGGCGCCCTG  1895

Qy  136  LeuLeuLeuGlyValArgAspLeuSerArgThrSerArgThrSerValTyrLeu---  154
Db 1896  CTGGAGCTTGGGGAGGAC--GCCAAGCTGCCCCCACGCTCACGGACTCTATGTCGGC  1952

Qy  155  LeuPheIleThrSerValLeuSerSerAlaProValAlaAspGlyProArgLeuGlnGly  174
Db 1953  CTGCTGGGCCGTGCAGTGCAGCCCTGCACAGCCCCCC--------------GGG  1988

Qy  175  AspLeuArgAsnLeuCysArgLeuAlaArgGluGlyValLeuGlyValArgArgAlaGlnPhe  194
Db 1989  GCCCTGGCAGAGCTGGCCAAGCTGCCTGGGAG-----CTGGGCCGCGACACATCAA--  2039

Qy  195  AlaGluLysGluGluGluGlnLnLeuGluLeuArgGlySerLysValGlnThrLeuPheLeu  214
Db 2040  -----AGTACCCTACAGGAGGACCAGTCCCATCCGCAGACGTGGAGGACCTGGGCGATG  2093

Qy  215  SerLys-------LysGluLeuProGlyValLeuGluThrGluValThrTyr---Gln  230
Db 2094  GCCAAAGGCTTAGTCCAACACACCCGCGGGCGCCAGCTGGCCTTCCCCAGC  2153

Qy  231  PheIleAspGlnSerPheGlnGluPheLeu--AlaAlaLeuSerTyrLeuGluLeuGluAsp  249
Db 2154  TTCCTCCTGCCAATGTTCCTGGGGGCCCCTGGCTGTGGCCTGAGTGGTGGCGAAATCAAGGAC  2213

Qy  250  GlyGlyValProArgThrAlaAlaGlyGlyValGlyThrLeuLeuArgGlyValAspAlaGln:  269
Db 2214  AAGGAGCTCCCGCAGTACCTAGCA-----------------TTGACCCCCAAGGAGAAGAGG  2258
```

```
Qy  412  ---------ArgLeuValAlaAlaGlnGluLys--------------LysLL  42:
Db 2820  ATGGGGGAGACCAAGCACTTCAGGAGGACCAGGAAGTTCACCACCGAGCCTTCAAAG  2879

Qy  421  ysLysSer-------LeuGlyLysArgLeuGlnLeuGlnAlaSerArgLeuGlyGlyG  435
Db 2880  CCAAGTCCCTGAAGGATGTGGAAGACCTGGAAGCTCCGGTGTCGGGGACCTAAAGAAACTGGAGT  2939

Qy  435  -ySerSerGlnGlyThrThrLysGlnLeuProAla-------------S  447
Db 2940  GTTCCTCGGAAGACACAGCAGCTGGGGAGCTCCCGCTGTCCGGGACCTAAAGAAACTGGAGT  2999

Qy  447  erLeuLeuHisProLeuPheGlnAlaMetThrAspPro-------------  459
Db 3000  TTGGGCTGGGCCCTGTCTCAGCCCCCCCAGGCTTCCCCCAAACTGTGCGGACTCCTCACGG  3059

Qy  460  -------LeuCysHisSerLeuSerLeuThrLeuSerHisCysLysLeuProAspA  476
Db 3060  CCTTTCCTCCTGCAGCATCTGACCTGGATGATGGCTGAGTGGAGAACAAGATGGGGACC  3119

Qy  476  laValCysArgAspLeuSerGluAlaAlaGlnAlaAlaLeuAlaProAlaLeuThrGluLeuGly:  496
Db 3120  AGGGTGTCTCGCAGGTCCACCCACCTTCCCACCAGCTGAGCCCTTCGAAACCCTCAATC  3179

Qy  496  euLeuHisAspArgLeuSerGluA-aGlyLeuArgMetLeuSerGluGlyValLeuAla-rpP  516
Db 3180  TGTCCCAGAACAACATCACCACCTGGGTGCCACAAACCTCGCCGAGGCCCTG------C  3233

Qy  516  roGlnCysArgValGlnThrValValGln-------------  526
Db 3234  CTTCGCTCGCTGCTCATCCCTCTCAGGCTAAGCTGTACGATAACTGCATCTCGCGACGTGG  3293

Qy  527  -----------LeuProAsp--------------------  529
Db 3294  GAGCCGAGAGCTTGGCTGCGTGTCGTCTTCCGGACATGGTGTCCCTCGGGTGATGGACGTCC  3353

Qy  530  ----------ProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSerP  545
Db 3354  AGTACAACAAGTTCACGGCTCCAGCCCTGGGGCCCAGTGCAGCCCTTCGGGAGGTGTC  3413

Qy  545  roAlaLeuThrThrLeuAspAspLeuSerGlyCysGlnLeuProCysAlaProMetValThrTyrL  565
Db 3414  CTCATGTGGAGAGCGGCTGGCGATGTGGACGGCCCACCATCAGTGTCCAGGAACACC  3473

Qy  565  eu-------------------CysAlaValLeuGlnHisGlnGlyC  574
Db 3474  TGCAACAACAGGATTCACGGACGATCAGCCTGAGATGACCCCAGTGTGCTCTCGGACAGGGCA  3533

Qy  574  ysGlyLeuGlnThrLeuSerLeuSerLeuAlaSerValGluLeuSerGluGlnLeuGlnLeuGlu--  593
Db 3534  GTTCCTGAGGACACTAACCACGCTGGACCTTGAACTGGGTACTTGTGGACACAGCTCTT  3593

Qy  594  LeuGlnAlaValLysArgAlaLysProAspAspLeuValIleThrHisProAlaLeuAspLeuA  613
Db 3594  CTCCAGGCTTGTATCCCATGAGGCCTCA------GCAtCCTTGGCACGACCGAGGGCCCTGCTGG-  3647

Qy  614  HisProGlnProProLysGluGlyLeu  621
Db 3648  TCAGGGTTGGCCCCTGCCCGGCTG  3671
```

```
RESULT 4
US-09-641-999-2
Sequence 2, Application US/09641999
Patent No. 6379894
; GENERAL INFORMATION:
APPLICANT: MACH, BERNARD
TITLE OF INVENTION: METHOD FOR SCREENING COMPOUNDS CAPABLE OF INHIBITING
TITLE OF INVENTION: FIXING BETWEEN THE STAT1 TRANSCRIPTION FACTOR AND THE
TITLE OF INVENTION: USF1 TRANSCRIPTION FACTOR
FILE REFERENCE: EGYP 3.3-007CONT
CURRENT APPLICATION NUMBER: US/09/641.999
CURRENT FILING DATE: 2000-08-18
```

```
Qy    526 lnLeuProAspProGlnArgGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSerProA 546
              ::||||||                                    ||||| ||
Db   3268 GGCTGCCG------------------------GGGCC-CAGAGCTGCTGCCAGCCTTCGGAGGTGTCCTC 3314

Qy    546 laLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyrLeu- 565
              ::: ||||||  ||| :::::  ||||||           ::: :|||||
Db   3315 ATGTGGAGACGGTGGCGATGTGGACGCCCACCATCCCATTCCAGTGTCCAGGAACACCTGC 3374

Qy    566 --------------------------------CysAlaValLeuLeuHisGlnHisGlyCysG 575
                                                 |||||||||||
Db   3375 AACAACAGGATTCACGGATCACGCCTGAGATGATCCCAGCTGTGCTCTGGACAGGCATGTT 3434

Qy    575 lyLeuGlnThrLeuSerLeuLeuAlaSerValGluSerGluGlnSerLeuGlnGlu-Leu 594
              :: :||||||                          ||||||
Db   3435 CTCTGAGGACACTAACCACGCTGGACCTTGAACTGGTACTTGTGGACACAGCTCTTCTC 3494

Qy    595 GlnAlaValLysArgAlaLysProAlaLeuValIleThrHisProAlaLeuAspGlyHis 614
              |||||||||                             ||||||
Db   3495 CAGGCTGTATCCCATGAGGCCTCA------GCATCCTGGCACCCGGCCCCTGCTGTTCA 3548

Qy    615 ProGlnProProLysGluLeu 621
              |||||||
Db   3549 GGGTTGGCCCCTGCCCGGCTG 3569
```

```
Alignment Scores:                              Length:              3393
Pred. No.:           2.09e-21                  Matches:              178
Score:               304.50                    Conservative:          78
Percent Similarity:  37.54%                    Mismatches:           263
Best Local Similarity: 26.10%                  Indels:               165
Query Match:         9.46%
```

```
Qy    270 ProHisSerHis---LeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGlu 288
              ||| :::  ::                    ||||||  ||||||:::
Db   2259 CCCATGACAACTGGCTGAGGGGTGCCACGGCTTTCTGGCTGGGCTGATCTTCCAGCCT 2318

Qy    289 ArgMetArgAspIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGlu 308
              ||| :: ||||| ||||||      ||                :|||::
Db   2319 CCCGCCCGCTGCCTGGGAGCCCTACTCGGGCCATCGGGCTGCCTCGGTGGA-CAGGAA 2377

Qy    309 AlaLeuArgTrpValGlnGlyGlnGlyGlyCys--ProGlyValAlaProGluValT 328
              :||  ||||||::                      ||||| :::
Db   2378 GCAGA-AGGTGCTTGCGAGGTACCTGAAGCGGCTGCAGCCGGGG--------------A 2421

Qy    328 hrGluGlyAlaLysGlyLeuIleAspThrGluGluProGluGluGluGluGluGlyGluG 348
              |||||| |||:::                  ||| ||:::
Db   2422 CACTGCGGGGCGGCCAGCTG----------------------------------- 2441

Qy    348 luProAsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheV 369
              ---------------CTTGAGCTGCTGCACTGCCCCACGAGCGCCGAGGACGCTGGAATTT 2487
Db   2442

Qy    368 alArgIleGlnAlaLeuCysArgPhePro---GluLeuAlaLeuGlnArgAlaValArgPheCysA 387
              :::                            |||
Db   2488 GGCAGCACGTGGTACAGGAGCTCCCCGGGCCCTC--CTTTCTGGGCACCCGCCTCACGC 2547

Qy    387 rgMetAspValAlaValLeuSerTyrCysValArgCys 399
              :::
Db   2548 CTCCTGATGCAACATGTACTGGCAAGGCCTGGAGGCGGCGGGCCAAGACTTCTCCCTGG 2607

Qy    400 -----------------CysProAlaGly------GlnAlaLeuArgLeuIleSerCys- 411
              :: :||                       |||||| |||
Db   2608 ACCTCCGCAGCACTGGCATTTGCCCCTCGGATTGGGAGCCCTGCTGGGACTCAGCTGTG 2667

Qy    411 ----------------------------------------------------------- 411
Db   2668 TCACCCGTTCAGGGCTGCCTTGAGCACAGGTGCCGCCGTGGAGTCCCTCGGGCAGC 2727

Qy    412 ----------------ArgLeuValAlaAlaGlnGlyLys-------------CysL 421
              :: :||
Db   2728 ATGGGGAGACCAAGCTACTTCAGGCAGCAGGAGAAGTTCACCATCGAGCCTTCAAAG 2787

Qy    421 ysLysSer------------LeuGlyLysArgLeuGlnAlaSerLeuGlyGly-Lys 435
              :: ||
Db   2788 CCAAGTCCCTGAAGGATGTGGAAGACCTGGAAAGCTTGTGCAGACTCAGAGGACGAGAA 2847

Qy    435 lySerSerGlnGlyThrThrLysGlnLeuProAla----------------------- 446
              |||||||| :::    ||||||
Db   2848 GTTCCTCGGAAGACACCAGCTGGGGAGCTCCTGCTGTCTCGGACCTAAAGAAACTGGAGC 2907

Qy    446 ----------------------------------------------------------- 446
Db   2908 TTGGCGCTGGGCCCTGTCTCAGGCCCCCCAGAGGCTTTCCCCAAACTGTGCGGATCCTCACGG 2367

Qy    447 --------SerLeuHisProLeuPheGlnAlaMetThrAsp------------------ 458
              ::::::
Db   2968 CCTTTTCCTCCCGCAGCATCTGGACCTGGATGCCGCTGAGTGAAGACAAGATGCGGGACG 3027

Qy    459 ----------------------------ProLeuCysHisLeuSerSerLeuThrL 468
              ||| |||||||
Db   3028 AGGGTGTCTCCGAGCTCTCAGCCACCTTCCCCAGCTGAAGTCCTTGGAAACCCTCCAATC 3087

Qy    468 euSerHisCysLysLysLeuProAspAlaValCysArgAspLeuSerGluAlaLeuArgAla- 487
              |||| ||||||||||||||
Db   3088 TGTCCCAGAACAACATCACTGACCTGGGTGCCTACAAACTCGCCAGGCCCCTGCCTTCGC 3147

Qy    488 --AlaProAlaLeuThrGluThrGluLeuGlyLeuHisAsnArgLeuSerGluGluAlaGlyLeuA 507
              ::::::
Db   3148 TCGCTGCATCCCGCTGCTCAGGCTAAGCTTGTACAATAACTGCATCTGCGACGTGGGAGCCG 3207

Qy    507 rgMetLeuSerGluGlyLeuAla---TrpProGlnCysArgValGlnThrValArgValG 526
              :::
Db   3208 AGAGCTTGGCTGCTGTCGTGTCTTCCGGACATGCGTCCCTCCGGGCGTGATGGACCGAAGTTCAC 3267
```

DB:

US-10-029-347-2 (1-625) x US-08-295-502-1 (1-3393)    Gaps: 27

```
Qy      1 MetLeuAlaGlnProGlyArgLeuLeuPheIleLeuAspGlyAlaAlaAspGluLeuProAla  20
            :::|||  |||||  ||||  |||||  |||| :::||||:|||
Db   1459 ATCTTGAAGAGACCTGACCCGGTTCTGCTCATCTGACGCCTTCTAGACGCCTCGAGGAGCTGGAAGCG 1518

Qy     21 LeuGlyGlyProGluAlaAlaProCys---ThrAspProPheGluAlaAlaSerGlyAla  39
            :::|| ||| ||  |||  ||| :::|||||  |||
Db   1519 CAAGATGGCTTCTGCACAGCCACGTGCGACCCGGACCGGCCGGAGCCCTGCTTCCCTCCGG 1578

Qy     40 ArgValLeuGlyLeuLeuSerLysLysAlaLeuLeuProThrAlaLeuLeuLeuValThr  59
            ||||| |||| |||| |||| ||||
Db   1579 GGGCTGCTGGCCGGCCTTTCCAAGAAGAGCTGCTCCGAGGTTGCACCCTCTCCTCACA 1638

Qy     60 ThrArgAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCysAla---  78
                                  CysProAlaGly--GlnAlaLeuArgLeuIleSerCys---
Db   1639 GCCCGCCCCGG-----GCCCGCCTGGTCCAGAGCCTGAGCAAGGCCGAGCCCCATTT 1692
                   ACCTCCGCAGCACTGGCATTGCCCCCCTGGATTGGGAGCCTCGTGGGACTCAGCTGTG 2644

Qy     79 GluValArgGlyPheSerArgAspLysAspGlyLysLysTyrPheTyrLysPheTyrLeu  99
            |||| ||||
Db   1693 GAGCTGTCGGCTTTCTCCATGAGACGCAGGCCCAGCATACGTGATGCCTACTTGAGCAGC 1752

Qy     99 GluLeuArgArgAla-------ArgAlaTyrArgPheValLysGluAsnGlnThrLeu 115
            |||                                 |||
Db   1753 TCAGGGATGACAGAGCACCGAGACCAGAGCCCTGAGCCCTGGGGAACCGGCCACTTCTT 1812

Qy    116 PheAlaLeuCysPheValProPheVa:CysTrpIleValCysArgGlnGln 135
            ||||                ::||||||| ||
Db   1813 CTCAGTCACAGCCACAGCCTACTTGTGCGGGAGTGTGCCAGCTCCAGAGGCCCTG 1872

Qy    136 LeuGluLeuGlyArgArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu-- 154
            ||||| ||||
Db   1873 CTGGAGCTTGGGGAGGAC--GCCAAGCTGCCCAGTGTGCCCACGCTTATGTCGGC 1929

Qy    155 LeuPheIleThrSerVa:LeuSerArgAlaProValAlaAspGlyProArgLeuG:nGly 174
            |||
Db   1930 CTCTGGGCCCGTCAGCCCGTCGACAGCCCCC-------GGG 1965

Qy    175 AspLeuArgAspAsnLeuCysArgGluAlaArgGluLeuValLeuGlyValLysArgArgAlaGlnPhe 194
            |||||                                         ||||||
Db   1966 GCCCTGGCAGAGCTGGCCAAGCTGGCCTGGGAG----CTGGCCCGAGACATCAA--- 2016

Qy    195 AlaGluLysGlyGluIleGlyGluArgGlyLeuArgGlySerLysVa:Gl:ThrLeuPheLeu 214
            ||||||| :||||| |||:||
Db   2017 AGGGTGTCTGCAGCTCTCAGCACCTTCCCCCAGCTGAGGAGACCTGGGCCCATG 2070

Qy    215 SerLys-------LysProGlyValLeuProGlyLeuThrGluValThrTyr---Gl:n 230
            :::|||
Db   2071 GCCAAAGCTTAGTCACACACCCCTGGAGGTCCAGAGTCCGAGCCTGGCCTCCCCAGC 2130

Qy    231 PheIleAspGlnSerPheGl:nGluPheLeu--AlaAlaLeuSerTyrLeuLeuGluAsp 249
            ||||
Db   2131 TTCCTCCTGCAATGCTTCCGGGGGCCCTGTGGCTGGCTC:CGAGTGGCGAAATCAAGGAC 2190

Qy    250 GlyGlyVal:ProArgThr:AlaLaG::yGlyVa:yGlyValAlaGlyThr:LeuLeuArgGlyAlaGl: 269
            |||                     |||
Db   2191 AAGGAGCTTCCCAGTACC~AGCA-----TTGACCCCAAGGATGGGAGAGAGG 2235

Qy    270 ProHisSerHis----LeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGl:n 286
            :::|||||| ||:|||
Db   2236 CCCATGACAACTGGCTGAGGGCGTGCACGCTTTCTGCTGGCTGCGATCTTCCAGCCT 2295

Qy    289 ArgMetArgAsplleGluArgHisPheGlyCysMetValSerGluAlaValLysGlnGlu 308
            :::|||                  |||||
Db   2296 CCCCCGCCTGCCTGGAGGGCCCTACTCGGCGGCCATCCGGGGCTGCCTCGGTGGA-CAGGAA 2354

Qy    309 AlaLeuArgTrpValGlnGlyGlnGlyGlyCys--ProGlyValAlaProGluValT 328
            :::|| :::||||                      |||||
Db   2355 GCAGA-AGGTGCTTGCGAGGTACCTGAAGCGGTCAGCCCGGG----------A 2398

Qy    328 hrGluGluGlyAlaLysGlyGlyLeuAspThrGluGluGluGluGluGlyGlyGluG 348
```

```
Qy   2399 CACTGCGGGCGCGGCAGCTG-------------  2418
            |||:::
Db    348 luProAsnTyrProGluGluLeuLeuTyrCysLeuTyrGluThrGlnGluAspAlaPheV 368

Qy   2419 ---CTGAGCTGTGCACTGGCCATGCGGCCACGAGGAGCCGAGGAGGCTGAATT 2464
            :::||||| ::||||
Db    368 aLArgGlnAlaLeuCysArgPro-- GluLeuAlaLeuGlnArgValArgPheCysA 387

Qy   2465 GGCGACGGTGGTACAGGAGCTCCCCGGCCGCCTCGTTTCTGGGCACCCGGCCTCCACGC 2524
            ||| :::
Db    387 rgMetAspValAlaValLeuSerTyrCysValArgCys------------------ 399

Qy   2525 CTCCGATGCCATGCTACTGGCCAAGGCCTGGACGCGGCGGGCCAAGACTTCTCCCTGG 2584
            :::
Db    400 --------CysProAlaGly--GlnAlaLeuArgLeuIleSerCys- 411

Qy   2585 ACCTCCGCAGCACTGGCATTGCCCCCTGGATTGGGAGCCTCGTGGGACTCAGCTGTG 2644

Qy   2645 TCACCCGTTTCAGGGCGTGCCTGACGCACGTGCGCGTGCTGTGGGAGTCCTGCGGCAGC 2704
            :::
Db    412 ------------ArgLeuValAlaAlaGl:nGluLys--------Lys 421

Qy   2705 ATGGGGGAGGACCAAGCTACTTCAGGCAGCAGGAGAAGTTCACCATCGAGCCTTCAAAG 2764
            :::|||
Db    421 ysLysSer----------LeuGlyLysLysArgLeuGlnAlaSerLeuGlyGlyG 435

Qy   2765 CCAAGTCCCTGAAGGATGTGAAGACCTGTGCAGACTCAGAGGACGAGAA 2824
            ||||
Db    435 lySerSerGlyThrThrLysLysGlnThrSerGlyGlyProArgLeuGly---------S 447

Qy   2825 GTTCCTGCGGAAGACACAGCTGGGGAGCTTCCCGTGTTCGGGACCTAAAGAAACTGGAGT 2884

Qy   2885 TTGGCTGGGCCCTGTCTCAGCCCCCCAGGCTTCCCCAAACTGGTGCGGGATCCTCACGG 2944
            :::
Db    447 erLeuLeuHisProLeuPheGlnAlaMetThrAspPro------------------ 459

Qy   2945 CCTTTTCCTCCTGCAGCATCTGGACCTGGAACAAGAATCGGGGACG 3004
            :::|||||
Db    460 --------LeuCysHisLeuSerSerLeuThrLeuSerHisCysLysLysProAlaAspA 476

Qy   3005 AGGGTGTCTGCAGCTCTCAGCCACCTTCCCCCAGCTGAAGCCCTCAATC 3064
            :::
Db    476 laValCysArgAspLeuSerGluAlaLeuLeuArgAlaAlaLeuThrGluGluG:yL 496

Qy   3065 TGTCCCAGACAACATCACTGACCTGGGTGCCTACAAACTCGCCGAGGCCCTG-----C 3118
            :::|||||
Db    496 euLeuHisAsnArgLeuSerGluAlaGlyLeuArgLeuMetLeuSerGl:uGlyLeuA:aTrpP 516

Qy   3119 CTTCGCTCGCTGCATCCCGTCCTCAGGCTAAGCTGTACACATAACTGCATCGTCGACGTGG 3178
            :::
Db    516 ------------------------------------LeuProAsp 526

Qy   3179 GAGCCCAGAGCTTGGCTCGTGCTTCGGACATGGTCCTCCGGGTGATGGACGTCC 3238
            :::|||||
Db    526 --------LeuProGlyLeuGlnTyrLeuValGlyMetLeuArgGlnSerP 545

Qy   3239 AGTACAACAACTGGCTGAGGGCGTGCACGCTTTCTGCTGGCAGCTTCGGAGGTGTC 3298
            :::
Db    545 roAlaLeuThrThrLeuAspLeuSerGlyCysGlnLeuProAlaProMetValThrTyrL 565

Qy   3299 CTCATGTGGAGACGGCTGGGCGATGTGGACGCCCACCATCCATTCAGTGTCCAGGAACACC 3358

Qy   3359 TG 3360
```

Db    565 eu 565

RESULT 6
PCT-US95-10691-1

; Sequence 1, Application: PCT/US95/10691
; GENERAL INFORMATION:
; APPLICANT: Glimcher, Laurie H.
; APPLICANT: Zhou, Hong
; APPLICANT: Douhan III, John
; TITLE OF INVENTION: METHODS OF IDENTIFYING COMPOUNDS
; TITLE OF INVENTION: USEFUL FOR TREATING AUTOIMMUNE DISEASES
; NUMBER OF SEQUENCES: 1
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Fish & Richardson
; STREET: 225 Franklin Street
; CITY: Boston
; STATE: MA
; COUNTRY: USA
; ZIP: 02110-2804
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.30B
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: PCT/US95/10691
; FILING DATE: 22 August 1995
; CLASSIFICATION:
; ATTORNEY/AGENT INFORMATION:
; NAME: Freeman, John W.
; REGISTRATION NUMBER: 29,066
; REFERENCE/DOCKET NUMBER: 00246/188001
; INFORMATION FOR SEQ ID NO: 1:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3393 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
; FEATURE:
; NAME/KEY: CDS
; LOCATION: 1..3393
PCT-US95-10691-1

Alignment Scores:
Pred. No.:
Score:                2.09e-2?
Percent Similarity:   304.50
Best Local Similarity: 37.54%
Query Match:          26.10%
DB:                   9.46%

Length:         3393
Matches:         178
Conservative:     78
Mismatches:      263
Indels:          165
Gaps:             27

US-10-029-347-2 (1-625) x PCT-US95-10691-1 (1-3393)

US-10-029-347-2 (1-625) x US-09-099-041A-9 (1-2859)

Qy    1  MetLeuAlaAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20

Db  814  CTGCTGCGCTTCCCCACGTGGCCCTTCACCTTCGATGGCGCTGCAGAGCTGCACTCG  873

Qy   21  ------LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSer  37

Db  874  GACTTGGACCTGAGCCGCGTGCCTGCAGCTCCTGC------CCCTGGAGCCTGCGCCAC  927

Qy   38  GlyAlaArgValLeuGlyyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeu  57

Db  928  CCCCTGGTCTTGTCGGCCAACCTGCTCAGTGGGAAGTGCTCAGCGGGCTAGCAAGCTG  987

Qy   58  ValThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuLeuCysSerProGlnCys  77

Db  988  CTCACAGCCCGCACA------GGCATCGAGTGCCGGCCAGTTCCTGCGGAGAAG  1038

Qy   78  AlaGluValArgGlyPheSerAspLysAspLysLysLysLysTyrPheTyrLysPhePheArg  97

Db 1039  GTGCTTCTCCGGGGCCTTCTCCCCAGCCACCTGGCGCTATGCCAGGAGGATGTTCCC  1098

Qy   98  AspGluArgArgAlaGlyArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAla  117

Db 1099  GAGCGGGGCCCTGCAGGACCGCCTGAGCCCAGTCGAGCCCAACCCAACCTCTGCAGC  1158

Qy  118  LeuCysPheValProPheValCysTrpIleValCysThrVal------LeuArgGln  134

Db 1159  CTGTGCTCTGTGCCCCTCTTCTGCTGGATCATCTTCCGGTGCTTCCAGCACTTCCGGTGCT  1218

Qy  135  GlnLeuGluLeuGlyValArgArgLeuSerArgThrSerArgThrThrSerValTyrLeu  154

Db 1219  GCCTTTGAAGGCTCACCAGCTGCCCGACTGCAGCATGCACGATGACCCTGACAGATGTCTTCCTC  1278

Qy  155  LeuPheIleThrSerValLeuSer------  162

Db 1279  CTG---GTCACTGAGGTCCA-TCTGAACAGGATGCAGCCCAGCCTGGTGCAGCGGAA  1334

Qy  163  ---SerAlaProValAlaAspGlyProArgLeuGlnGly-AspLeuArgLeuAsnLeuCysArg  181

Db 1335  CACACGCCAGCCCAGTGGAGACCCTCCACGCGCCGGGACACTCTGTGCTGCTGCTGGGGCA  1394

Qy  181  gLeuAlaArgGluGluGlyValLeuValLeuArgArgAlaGlnPheAlaGlnLeuLysGluGluGlu  201

Db 1395  GGTGGGCCCACCCGGGCA:GGAGAAGAGCCTCTTTGTCTTCACCCAGGAGGAGGTGCAGGC  1454

Qy  201  nLeuGluArgGlySerLysValGlnThrLeuPheLeuSerLysLysGluPro--  220

Db 1455  CTCCGGGGCTGCAGGAGAGACATGCAGCTCGGGCTTCCTG-----CGGGCTTTGCCGGA  1508

Qy  221  ---GlyValLeuGluThrGluValThrValTyrGlnPheIleAspGlnSerPheGlnGluPh  239

Db 1509  GCTGGGCCCCGGGGTGACCAGCAGTCCTA:GAGTTTTCCACCTCACCCCCCAGGCCTT  1568

Qy  239  eLeuAlaAlaSerTyrLeuGluAsp-------  249

Db 1569  CTTTACAGCCTTCTCCTCCTGGTGCTGGACGAGGGTGGGCACTCAGGAGTCTGCTCAGGTT  1628

Qy  250  -------GlyGlyValProArgThrAlaAla  257

Db 1629  CTTCCAGGAGTGGATGCCCCTGCGGGGGCAGCGACCCAGTCCGCTCATCCCCCTTCCT  1688

Qy  258  --------------GlyGlyValGlyThrLeuLeuArgGlyAspAlaAlaGlnProHisSe  272

Db 1689  CCCGTTCCAGTGCCTGCAGGCGAGCAGTGTCCGGCGCGGGGAAGAACCTCTTCAAGAACAAGGA  1748

Qy  272  rHisLeuValLeuThrArgPheLeuPheGlyLeuLeuSerValAlaGluArgMetArgAs  292

Db 1749  TCACTTCAGTTCACCAACCTCTTCCTGTGGGGCTGTTGTCCAAAGCCAAACAGAA--  1806

Qy  292  pIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluAlaLeuArgTr  312

Db 1807  -CCCCTGCGGCATCGTGTGCCCCGGGCAGCCCTGAGGAGAAAGGCCAAGGCCCCTG---TG  1862

Qy  421  ysGlySer--------------LeuGlyLysArgLeuGlnAlaSerLeuGlyGlyG  435

Db 2765  CCAAGTCCCTGAAGGATGGTGGAAGACCTTGGTGCAGACTTCGCGAGGACGAGGAA  2824

Qy  435  lySerSerGlnGlyThrLysGlnLeuProA--------  447

Db 2825  GTTCCTCGGAAGACAGCTGGGGAGCTTCCTGTTCGGGAGCTCCCTGCGGTTCCCTGTTCGGGACCTGAAAGAAACTGGAGT  2884

Qy  447  erLeuLeuHisProLeuPheGlnAlaMetThrAspPro--------  459

Db 2885  TTGCGGTGGGCCGTGTCTCAGGCCCCAGGCTTCCCCAAACTCGGCGGATCCTCACGG  2944

Qy  460  --------LeuCysHisSerLeuSerLeuThrLeuSerHisCysSerLeuProCysAspA  476

Db 2945  CCTTTTCCTCCGCAGCATCTGGACCTGAGTGCCAGTGAGACCAAGATCGGGGACG  3004

Qy  476  laValCysArgAspLeuSerGluAlaLeuArgAlaAlaProAlaAlaLeuArgThrGlyLeuGly:  496

Db 3005  CTCCAAGCCCGGCACA------GGCATCGAGTGCCTTGGAAACCCTCAATC  3064

Qy  496  euLeuHisAsnArgLeuSerGluAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpP  516

Db 3065  TGTCCCAGAACAACATCACTGACCTGGGTGCCTACAAACTCGCCGAGCCCCTG------C  3118

Qy  516  roGlnCysArgValGlnThrValArgValGln--------  526

Db 3119  CTTCGCTCGTCGCTGACATCCCTGGTCCAGGTCAAGCTGTGTACAATAACTGCATCTGCGACGTGG  3178

Qy  527  --------LeuProAsp--------  523

Db 3179  GAGCCGAGAGCTTGGCTCGTGTGCTTCCGGCACATGGTGGTCCCCGGGTGATGGACGTCC  3238

Qy  530  ----ProGlnArgGlyLeuGlnTyr:LeuValGlyMetLeuArgGlnSerP  545

Db 3239  AGTACACAACAAGTTCACGGCTGCCGGGCTGCCCAGCTGCCTGCCAGCCTTCGGAGGTGTC  3298

Qy  545  roAlaLeuThrLeuAspLeuProLeuSerGlyCysGlnLeuProAlaProMetValThrTyrL  565

Db 3299  CTCCATGTGGAGACGCTGGCGATGTGGACGCCCCACCATCCATTCAGTGTCCAGGAACACC  3358

Qy  565  eu  565

Db 3359  TG  3360

RESULT 7
US-09-099-041A-9
; Sequence 9, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-076001
; CURRENT APPLICATION NUMBER: US/09/099,041A
; CURRENT FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-099-041A-9

Alignment Scores:
Pred. NO.:                                    Length:        2859
Score:                  289.00               Matches:        172
Percent Similarity:     37.84%               Conservative:   94
Best Local Similarity:  24.47%               Mismatches:     261
Query Match:                                 Indels:         158
DB:                                          Gaps:           27

Score:  6.65e-20

RESULT 8
US-09-245-281-9
; Sequence 9, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
;   TITLE OF INVENTION: AND USES THEREOF
;   FILE REFERENCE: 07334/118001
;   CURRENT APPLICATION NUMBER: US/09/245,281
;   CURRENT FILING DATE: 1999-02-05
;   EARLIER APPLICATION NUMBER: US 09/207,359
;   EARLIER FILING DATE: 1998-12-08
;   EARLIER APPLICATION NUMBER: US 09/099,041
;   EARLIER FILING DATE: 1998-06-17
;   EARLIER APPLICATION NUMBER: US 09/019,942
;   EARLIER FILING DATE: 1998-02-06
;   NUMBER OF SEQ ID NOS: 44
;   SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 9
;   LENGTH: 2859
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-245-281-9

Alignment Scores:
Pred. No.:                                    Length:        2859
Score:                  6.65e-20             Matches:        172
Percent Similarity:     289.00               Conservative:    94
Best Local Similarity:  37.84%               Mismatches:     281
Query Match:            24.47%               Indels:         158
DB:                      8.98%               Gaps:            27
                            4

US-10-029-347-2 (1-625) x US-09-245-281-9 (1-2859)

Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheLeuLeuLeuPheLeuLeuAspGlyAlaAspGluLeuProAla 20
            :::::::|||                  ||||||         |||
Db    814 CCGCTGCGCTTCCCCCAGGTGCCCTCTTCACCTTCGATGGCCTGGACGGCTGCACTCG 873

Qy     21 ------LeuGlyGlyProGluAlaAlaaProCysThrAspProPheGluAlaAlaSer 37
                                        :::  |||          :::::::
Db    874 GACTTGGACCTGCGCCGGTGCCTGCACAGCTCCTGC----CCTGGGAGCCCTGCCCAC 927

Qy     38 GlyAlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeu 57
            |||            :::              :::   ||      ||
Db    928 CCCCTGCTTCTTGCTGGCCAACCTGCTCAGTGGGAAGCTGCTCAAGGGGCTAGCAAGCTG 987

Qy     58 ValThrThrArgAlaAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCys 77
            :::::::|||                            ||
Db    988 CTCACAGCCCGCCACA------GGCATCGAGGTCCCGGCCAGTTCCTGCGGAAGAAG 1038

Qy     78 AlaGluValArgGlyPheSerAspLysAspLysLysLysTyrPheLysPhePheArg 97
            ::::||||||                       |||
Db   1039 GTGCTTCTCCGGGGCTTCTCCCCCAGCCACCTGCGGCGCCTATGCCAGGAGGATGTCCCCC 1098

Qy     98 AspGluArgArgAlaGlyAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAla 117
            :::                                          :::::::
Db   1099 GAGCGGGACCCTGCAGGACCGCCTGCTGCAGGCCAGCTGGAGGCCAACCCCACCTCTGCAGC 1158

Qy    118 LeuCysPhePheValProPheValCysTrpIleValCysThrVal------LeuArgGln 134
            |||||||                                            |||
Db   1159 CTGTGCTTCGTGCCCTTCGTGCTCTGCTGGATCATCTTCCGGTGCTTCCGGCACACTTCCGTGCT 1218

Qy    135 GlnLeuGluLeuGlyArgAspLeuSerArgLysSerThrArgLysThrThrSerValTyrLeu 154
            |||||||                    :::::::
Db   1219 GCCTTTGAAGGCTACCACCAGGCCACAGATGAGTGGCAGA 1278

Qy    155 LeuPheIleThrSerValLeuSer------ValIleThrHisProAl 162
            ||||||
Db   1279 CTG---GTCACTGAGGTCCA-TCTGAACACAGGAGTGCAGCCACCTGGTCGCAGCGGAA 1334

Qy    163 ---SerAlaProValAlaAspGlyArgLeuGlnGly·AspLeuArgArgAsnLeuCysArg 181
            :::   |||                            ||

```
Qy   503  uAlaAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrVa  523
          ||| :::::::     |||              :::|||
Db  2310  TGTCGGAGCCAGGCACGCACCCAAAAATCTG------GATGAATGCAAAGGCCTCACG--  2364

Qy   523  lArgValGlnLeuProAsp----------ProGlnArgGlyLeuGlnTyrLeu----------  537
          |||                       |||||
Db  2362  -CACCTTAAACTGGGAAAAACAAAATAACAAGTGAAGGAGGAGTATCTCGCCCTGGC  2420

Qy   538  ------------------------------------ValGlyMet-------------  540
                                              |||||||
Db  2421  TGTGAAGAACAGCAAATCAATCTCTGAGGTTGGGGCAATCAAGTGGGGGATGA  2480

Qy   541  --------------------------LeuArgGlnSerProAlaLeuThrThrLeuAspLe  552
                                    ::|||:||:|||||
Db  2481  AGGAGCAAAAGCCTTCGCAGAGGCTCTGCGGGAACCACCCCAGCTTGACCACCCTGAGTCT  2540

Qy   552  uSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysValAlaValGlnHisGl  572
          |||                                   :::|||
Db  2541  TGGTCCAACGGCATCTCCACAGAAGGAGAAGAGCCTTGCGAGGGCCTGCAG---CA  2597

Qy   572  nGlyCysGlyLeuGlnThrLeuAlaSerValGluLeuSerGluGl-nSerLeuGl  592
          |||::::|||
Db  2598  GAACACGTCTTCTAGAAATACTGTGGCTGACCCAAAATGAACTCAACGATGAAGTGGCAGA  2657

Qy   592  nGluLeuGlnAlaValLysArgAlaLysProAspLeu------ValIleThrHisProAl  610
          :::|||
Db  2658  GAGTTGGCAGAAATGTT-GAAGTCAACCAGACGTTAAAGCATTTATGGCTYATCCAGA  2716

Qy   610  aLeuAsp  612
          |||
Db  2717  ATCAGAT  2723
```

RESULT 9
US-09-207-359B-9
; Sequence 9, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John.
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-207-359B-9

Alignment Scores:
```
Pred. No.:                                    Length:          2859
Score:                   6.65e-20             Matches:          172
Percent Similarity:      37.84%               Conservative:      94
Best Local Similarity:   24.47%               Mismatches:       281
Query Match:             8.99%                Indels:           158
DB:                           4               Gaps:              27
```

US-10-029-347-2 (1-625) x US-09-207-359B-9 (1-2859)

```
Qy     1  MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuGluAspGlyAlaAspGluProAla  20
          :::|||                     |||||
Db   814  CTGCTGCGCTTCCCCCAGGTGGCCCTCTTCACCTTCGATGGCCTGGACGAGCTGCACTCG  873

Qy    21  ------LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSer  37
                                     ::|||
Db   874  GACTTGGACCTGAGCCGCCTGCCCTGACAGCTCCTGC------CCCTGGGAGCCTGCCCAC  927
```

(left column — continuation of prior alignments)

```
Db  1335  CACACGCAGCCCAGTGGAGCCCTCCACGCCGCCGGCGACCTCTGTGCTCGCTGGGGCA  1394
Qy   181  gLeuAlaArgGluGlyValGlyValArgArgAlaGlnPheAlaGluLeuLysGluLeuGluGl  201
Db  1395  GGTGGCCCACCGGGCATGGAAGAGAGAGCCTCTTGTCTTCACCCAGGAGGAGTGCAGGC  1454
Qy   201  nLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSerCysLysLysGluLeuPro-  220
Db  1455  CTCCGGGCTGCAGGAGAGACAATGCAGTGGGCTTCCTG-----CGGGCTTTGCCGGA  1508
Qy   221  ---GlyValLeuGluThrGluValThrGlyArgGlnSerPheGlnGlnGluPH  239
Db  1509  GCTGGCCCGGGG-GACCAGCAGTCCTATGAGTTTTCCACCCACCCTCCAGGCCT-  1568
Qy   239  eLeuAlaAlaLeuSerTyrLeuLeuGluAsp----------  249
Db  1569  CTTTACAGCCCTTCTTCCTCGTGCTGGACGACCAGGTGGGCACTCAGGAGCTGCTCAGGT  1628
Qy   250  p-------------GlyGlyValProArgThrAlaAla-----  257
Db  1629  CTTCCAGGAGGTGGATGCCCCCTGCGGGGGCAGCGACCAGTCGTCTGCTATCTCCTTCCT  1688
Qy   258  -------GlyGlyValGlyThrLeuLeuArgGlyAspAlaGlnProHisSe  272
Db  1689  CCCGTTCCAGTGCCTGCAGGCAGTGGTCCGGCGGGAAGACCTCTTCAAGAACAAGGA  1748
Qy   272  rHisLeuValLeuThrThrArgPheGlyLeuLeuSerAlaGluArgMetArgAs  292
Db  1749  TCACTTCCAGTTCACCAACCTCTTCCTGTGCGGGGCTGTTGTCCAAAGCCAAACAGAAA-  1806
Qy   292  pIleGluArgHisPheGlyCysMetValSerGluValLysGlnValCysProGlyValAlaProGl  312
Db  1807  -CTCCTGCGGCATCTGGTGCCCGCGGCCTGAGGAGAAAGCCAAGGCCCTG----TG  1862
Qy   312  p---------ValGlnGlyGlnGlyGlyCysProGlyValA-aProGl  326
Db  1863  GGCACACTGTTTCCAGCCTGCGGGGCTACCTGAAGAGCCTGCCCGCTTCAGTCGA  1922
Qy   326  uValThrGluGlyGlyAlaAlaLysGlyGluAspThrGluGluGluGluGluGluGluGluGlu  346
Db  1923  AAGCTTCAACCAGGTGCAGGCCATG-----  1947
Qy   346  yGluGluProAsnTyrProLeuGluLeuLeuTyrCysLeuTyrArgGluThrGlnGluAsp-  365
Db  1948  -------CCCACGTTC--ATCTGGATGGTCGCGCTCGATCCTCACGAGACACGAGCCAGA  1997
Qy   366  -------AlaPheValArgGlnAlaLeuCysAlaArgPheProGluLeuAlaLeuGlnLeuGlnArgVa  383
Db  1998  GGTGGGGCAGCTGGCGCGGCGCCATCTGGCCC---AACACCTCAAGCTGACCTACTG  2054
Qy   383  lArgPheCysArgMetAspArgValAlaAlaValSerValTyrCysValCysProAlaGl  403
Db  2055  CAAGCCCTGCTCGCCGACTGCAGCGCGCCCTCCTCCGTCCCTGCATCACTTCCCCAAGCCG  2114
Qy   403  yGlnAlaLeuArgLeuLeuSerCysArgLeuValA-aAlaGlnGluLysLysLysLysLysSe  423
Db  2115  GCTGGCCCTAGACCTAGACAACAACAATCTC-----  2145
Qy   423  rLeuGlyLysArgLeuGlnAlaSerLeuGlyLysGlyLysSerSerGlnGlyTyrThrThrLysGl  443
Db  2146  --------AACGACTACGGCGTGCGGGAG--  2166
Qy   443  nLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuProLeuCysHisLe  463
Db  2167  ------------CTGCAGCCCTGGCTTCAGCCGGCCTCACT-----  2193
Qy   463  uSerSerLeuThrLeuSerHisCysLeuLysLeuProAspAspAlaValCysArgAspLeuSerGl  483
Db  2194  ----GTTCTCAGATCCAGCTCGTTAAACCAGATCACCGACGGTGGGGTAAAGGTGCTAAGCGA  2249
Qy   483  uAlaLeuArgAlaAlaProAlaLeuThrArgGluLeuAspLeuLeuHisLeuAsnArgLeuSerGl  503
Db  2250  AGAGCTGGACCTGAGCCGCGTGCCCTGACAGCTCCTGC---  2309
```

Qy    346 yGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluAsp-- 365
          ||||||||   :::|||   :::||||||||||||||||||||||||||||||||
Db   1948 ------CCCACGTTC--ATCTGGATGCTGCGCTGCATTCGAGACACAGAGCCAGAA 1997

Qy    366 ------AlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArgVa 383
          |||   ||||||||||||||||||||||||||
Db   1998 GGTGGGGCAGGTGGCGCGGCCCAGGGCATCTGCGCC--AACTACCTCAAGCTGACCTACTG 2054

Qy    383 lArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysLysProCysAlaGl 403
          :|||||||||||||||||||||||||||||||||||||||||||||
Db   2055 CAACGCCTGCTCGGCCCACTGCAGCGCCGCCCTTCCTTCCGTCCTGCATCACTTCCCCAAGCG 2114

Qy    403 yGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnLeuLysLysLysSe 423
          |||||||  |||||    :::||
Db   2115 GCTGGCCCTAGACCTGGAGCCAGCTGGAGGCCAACCCCAACAACAATCTC-- 2145

Qy    423 rLeuGlyLysArgLeuGlyLysAlaSerLeuGlyGlySerGlnGlyThrThrLysGl 443
          :::||||||||||
Db   2146 ------AACGACTACGGCGGTGCGGGAG-- 2166

Qy    443 nLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLe 463
          |||||||||
Db   2167 ------CTGCAGCCCTGCTTCAGCCGCCTCACT-- 2193

Qy    463 uSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAspLeuSerGl 483
          |||  :::   |||
Db   2194 ------GTTCTCAGACTCAGCGTAAACCAGATCACACTGACGCGGTGGGGTAAAGGTGCTAGCGA 2249

Qy    483 uAlaLeuArgAlaAlaProAlaLeuThrGluLeuLeuHisAsnArgLeuSerGl 503
          |||||||||
Db   2250 AGAGCTGACCAAATACAAAATTGTGACCTATTTGGGTTTATCACACCAGATCACCGA 2309

Qy    503 uAlaGlyLeuArgMetLeuSerGluGlyGlyLeuAlaTrpProGlnCysAlaArgValGlnThrVa 523
          :::|||||||||
Db   2310 TGTCGGAGCCAGGTACGTCACCAAAATCCTG--GATGAATGCAAAGGCCTCACG-- 2361

Qy    523 lArgValGlnLeuProAsp-------ProGlnArgGlyLeuGlnTyrLeu------ 537
          |||||||
Db   2362 -CATCTTAAACTGGGAAAAAACAAAATAACAAGTGAAGGAGGAAGTATCTCGCCCTGGC 2420

Qy    538 ------------ValGlyMet----------------- 540
          |||||||
Db   2421 TGTGAAGAACAGCAAATCAATCTCGAGGTTGGGATGTGGGCAATCAAGTGGGGATGA 2490

Qy    541 ------LeuArgGlnSerProAlaLeuThrThrLeuAspLe 552
          |||||||||
Db   2481 AGGAGAAAAGGCCTTCGCAGAGGCTCTGCGGAACCACCCAGCTTGACCCCTGAGTCT 2540

Qy    552 userGlyGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGl 572
          :::
Db   2541 TGCGTCCAACGGCATCTCCACAGAGGAGAAAGAGCCTTGCGAGGGCCCTCAG--CA 2597

Qy    572 nGlyCysGlyLeuGlnThrLeuSerLeuAlaSerValLeuSerGluGlnSerLeuGl 592
          |||||
Db   2598 GAACACGTCTCAGAAATACTGTGGCTGGACCCAAAATGAACTCAACGATGAGTGGCAGA 2657

Qy    592 nGluLeuGlnAlaValLysArgAlaLysProAspLeu------ValIleThrHisProAl 610
          :::
Db   2658 GAGTTGGCAGAAATGTT-GAAAGTCAACCAGACGTTAAAGCATTTATGGCTTATCCAGA 2716

Qy    610 aLeuAsp 612
Db   2717 ATCAGAT 2723

RESULT 10
US-09-340-620A-9
; Sequence 9, Application US/09340620A
; Patent No. 6482933
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY AND USES THE
; FILE REFERENCE: 07334-124001
; CURRENT APPLICATION NUMBER: US/09/340,620A

Qy     38 GlyAlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeu 57
          |||||||||
Db    928 CCCCTGGTCTTGCTGCCAACCTGCTCAGTGGGGAAGCTGCGCAGTGGGGCTAGCAAGCTG 987

Qy     58 ValThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCys 77
          |||||||||
Db    988 CTCCACAGCCCGCACA------GGCATCGAGGTCCCGCGCCAGTTCCTGCGGAAGAAG 1038

Qy     78 AlaGluValArgGlyPheSerAspLysLysTyrPheTyrLysPhePheArg 97
          :::||||||||||
Db   1039 GTGCTTCTCCGGGGGTTCTCCCCAGCCACCTGCGCCTATGCCAGGAGGATGTTCCCC 1098

Qy     98 AspGluArgArgAlaGluArgAlaTyrArgAlaPheValLysGlyAsnGluThrLeuPheAla 117
          ::: |||  |||
Db   1099 GAGCGGCGGCCCTGCAGGACCGCCGCTGCTGAGCCAGCTGGAGGCCAACCCCAACCTCTGCAGC 1158

Qy    118 LeuCysPheValProPheValCysTrpIleValCysThrVal------LeuArgGln 134
          :::   |||
Db   1159 CTGTGCCCTGTGCCCCTCTTCGTGCTGGATCATCTTCCGGTGCTTCCAGCCACTTCCGTGCT 1218

Qy    135 GlnLeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu 154
          :::
Db   1219 GCCCTTTGAAGGCTCACCACACGCTGCCCGACTGACCCTGACACGATGTCTTCCTC 1278

Qy    155 LeuPheIleThrSerValLeuSer 162
          |||
Db   1279 CTG--GTCACTGAGGTCCA-TCTGAACAGGATGCAGCCCAGCCTGGTGCAGCGGA 1334

Qy    163 --SerAlaProValAlaAspGlyProArgLeuGlnGlyLy-AspLeuArgAsnLeuCysSerAr 181
          :::||||||
Db   1335 CACACCCAGCCCAGTGGAGACCCTCCACGCGCCCGGCACTCGTGCTGCGGTGGGGCA 1394

Qy    181 gLeuAlaArgGluGlyValHisGlyValArgArgAlaGlnPheAlaGlnLysGluLeuGluGl- 201
          :::
Db   1395 GGTGGGCCACCACCGGGCATGGAGAAGACCCTCTTGTCTTCACCCAGGAGGAGGTGCAGGC 1454

Qy    201 nLeuGlyValSerLysValGlnThrLeuPheLeuSerLysLysGluProPro-- 220
          :::   |||
Db   1455 CTCCGGGGCTGAGGAGAGACAATGGGGCTGGGCTTCCTG------CGGGCTTGCCCAGA 1508

Qy    221 ---GlyValLeuGluThrGluValThrTyrGlnPheIleAspGlnSerPheGlnGlnProh 239
          |||||||||
Db   1509 GCTGGGCCCCGGGGGTGACCAGCAGTTCTTCCACCTCACCCTCAGGCCTT 1558

Qy    239 eLeuAlaAlaLeuSerTyrLeuLeuGluAsp 249
          |||||||
Db   1559 CTTTACAGCCTTCTTCCTCG-GCTGGACGACAGGTGGGCACTTCAGGAGCTGCTCAGGTT 1628

Qy    250 ----------GlyGlyValProArgThrAlaAla---- 257
          |||||
Db   1629 CTTCCAGGAGTGATGCCCCTGCGGGGGGCAGCGACCACGTCCTGCTATCCTCCCTTCCT 1688

Qy    258 ------------GlyGlyValGlyThrLeuArgLeuArgGlnAlaGlnProHisSe 272
          :::
Db   1689 CCCGTTCCAGTGCTCCTGCAGGGCGTCCCGCGGCGTCCCGAGGGAGAACAAGGA 1748

Qy    272 rHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAs 292
          ||||||
Db   1749 TCACTTCCAGTTCACCACAACCCCTTCCTTCCGTGCGGGTGTTGTGCCAAAGCAGAAA-- 1806

Qy    292 pIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluGluAlaLeuArgTr 312
          |||||||   :::||||||
Db   1807 -CTCCTGCGGCATCTGGTGCCCGGGCCCAGCCCGGAGGAGAAGGCCAAGGCCCTG---TG 1862

Qy    312 p------------ValGlnGlyGlnGlyGlyCysProGlyValAlaProGl 326
          |||||||
Db   1863 GGCACACCTGTTTTCACCCGTGCGGGGTACCTGCAGAGCCTGAAGAGCCCGGTTCAGGTCGA 1922

Qy    326 uValThrGlyAlaLysGlyLeuAspThrGluProGluProGluGlyGluGlu-Gl 346
          :::   |||
Db   1923 AAGCTTCAACCAGGTGCAGGCCATG------ 1947

; CURRENT FILING DATE: 1999-06-28
; PRIOR APPLICATION NUMBER: US 09/245,281
; PRIOR FILING DATE: 1999-02-05
; PRIOR APPLICATION NUMBER: US 09/207,359
; PRIOR FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1999-02-06
; NUMBER OF SEQ ID NOS: 71
; SEQ ID NO 9
; LENGTH: 2859
; TYPE: DNA
; ORGANISM: Homo sapiens
US-09-340-620A-9

Alignment Scores:
Pred. No.:
Score:                   289.00
Percent Similarity:      37.84%
Best Local Similarity:   24.47%
Query Match:             8.95%

; SOFTWARE: FastSEQ for Windows Version 4.0

Length:        2859
Matches:       172
Conservative:  94
Mismatches:    281
Indels:        158
Gaps:          27

6.65e-29

US-10-029-347-2 (1-625; x US-09-340-620A-9 (1-2859)

Qy    1  MetLeuAlaGlnProGlnArgLeuPheIleLeuAspGluLeuAspGluLeuAlaProAla  20
Db  814  CTGCTGGCGCTTCCCCACGTGGCCCTTCACCTTGATGGCCTGGACGAGCTGCACTCG  873

Qy   21  ------LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSer  37
Db  874  GACTTGGACCTGACCAGCCGCGTGCCTGACAGCTCTGC------CCCTGGGAGCTGCCCAC  927

Qy   38  GlyAlaArgValLeuGlyValGlyLeuLeuSerLysLysAlaLeuProThrAlaLeuLeu  57
Db  928  CCCCTGGTCTTGTCTGGCCAACCTGCTTCTCAGTGGGAAGCTGCTAGCAGGGCTGCC  987

Qy   58  ValThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCys  77
Db  988  CTCACAGGCCCGGCACA------GGCATGGAGGTCCCGGCCAGTTCCTGCGGAAGAAG  1036

Qy   78  AlaGluValArgGlyPheSerAspLysLysLysTyrPheTyrLysLysPhePheArg  97
Db 1039  GTGCTTCTCCGGGGCTTCTCCCCCAGCCACCTGCCGCCTATGCCAGGAGGTGTCCCC  1098

Qy   98  AspGluArgArgAlaGlyAlaTyrArgArgPheValLysGluAsnGluThrLeuPheAla  117
Db 1099  GAGCGGCGGCCCTGCAGGACCGCCTGCTGAGGCCAGCTGGAGGCCAACCCCAACCTCTGCAGC  1158

Qy  118  LeuCysPheValProPheValCysTrpIleValCysThrVal------LeuArgGln  134
Db 1159  CTGTTGCTGTGCCCCCTTCTGTCTGGATCATCTTCCGGTGCACTTCCGTGCT  1218

Qy  135  GlnLeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu  154
Db 1219  GCCTTTGAAGGCTCACCACAGTGCCCGACTGCACGATGCCTGACAGATGCCTTCCTC  1278

Qy  155  LeuPheIleThrSerValLeuSer------  162
Db 1279  CTG--GTCACTGAGGTCCA-TCTGAACAGGATGGCAGCCAGCCAGCGGCAGGGAGCGGGAA  1334

Qy  163  ---SerAlaProValAlaAspGlyProArgLeuGlnGly-AspLeuArgAsnLeuCysAr  181
Db 1335  CACACGAGCCAGCCCAGTGGAGACCCCTGCCAGCCGCGGCGGACACCTGTGCTCCGCTGGGGCA  1394

Qy  181  GlnAlaGlyLeuArgMetLeuSerGlyGlyLeuLeuAlaGlnPheA-aGlyLysGluGlu-  201
Db 1395  GGTGGCCCCACCGGGCATGGAGAGAAGGCCTCTTTGTCTTCACCCAGGAGGAGGTGCAGGC  1454

Qy  201  nLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeuSerLysLysGluGluLys  220
Db 1455  -CATCTTAAACTGGGAAAAAAACAAAATAACAAGTGAAGGAGGAAGTATCTCGCCCTGGC  2420

Qy  221  ---GlyValLeuGluThrGlu...

Db 1455  CTCCGGGCTGCAGGAGAGAGACATCAGCTGGGCTTCCTG-------CGGGCTTTGCCGGA  1508

```
Qy    538 ------------------------------ValGlyMet----------------- 540
                                       ||||||
Db   2421 TGTGAAGAACAGCAAATCAATCTCTGAGGTTGGGATGTGGGCAATCAAGTTGGGA-GA 2480

Qy    541 ---------------LeuArgGlnSerProAlaLeuThrThrLeuAspLeu 552
                                ||||||
Db   2481 AGGAGCAAAAGCCTTCGCAGAGGCTCTGCGGAACCACCCCAGCTTGACCACCCTGAGTCT 2540

Qy    552 uSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValLeuGlnHisGl 572
              :::             ||||              ||||||
Db   2541 TGCGTCCAACGGCATCTCCACAGAGGAGGAAAAGAGCCTTGGCGAGGGCCCTGCAG---CA 2597

Qy    572 nGlyLeuGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGl 592
                  |||||
Db   2598 GAACACGTCTTCAGAAATACTGTGGCTGACCCAAAATGAACTCAACGATGAAGTGGCAGA 2657

Qy    592 nGluLeuGlnAlaValLysArgAlaLysProAspLeu--------ValIleThrHisProAl 610
                       |||||   ||||||
Db   2658 GAGTTTGGCAGAAATGTT-GAAGTCAACCAGACGTTAAAGCATTATGGCTTATCCAGA 2716

Qy    610 aLeuAsp 612
          |||
Db   2717 ATCAGAT 2723
```

RESULT 12
US-09-099-041A-7
; Sequence 7, Application US/09099041A
; Patent No. 6340576
; GENERAL INFORMATION:
;   APPLICANT: Bertin, John
;   TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
;   TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
;   FILE REFERENCE: 07334-076001
;   CURRENT APPLICATION NUMBER: US/09/099,041A
;   CURRENT FILING DATE: 1998-06-17
;   PRIOR APPLICATION NUMBER: 09/019,942
;   PRIOR FILING DATE: 1998-02-06
;   NUMBER OF SEQ ID NOS: 37
;   SEQ ID NO 7
;   LENGTH: 3392
;   TYPE: DNA
;   ORGANISM: Homo sapiens
;   FEATURE:
;   NAME/KEY: CDS
;   LOCATION: (245)...(3103)
US-09-099-041A-7

Alignment Scores:
Pred. No.:
Score:                     8.8e-20
Percent Similarity:        289.00
Best Local Similarity:     37.84%
Query Match:               24.47%
                            9.98%
DB:

Length:        3392
Matches:        172
Conservative:    94
Mismatches:     281
Indels:         158
Gaps:            27

US-10-029-347-2 (1-625) x US-09-099-041A-7 (1-3382)

```
Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla 20
             :::||      |||||          ||||||            :::
Db   1058 CTGCTGCGCTTCCCCAGTGGCCCTTCACCTTCGATGGCCTGACGAGCTGCACTCG 1117

Qy     21 --------LeuGlyTyrGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSer 37
                    ||||                            |||
Db   1118 GACTTGGACCTGAGCCGCCGTCCTGCCTGACAGCTTCCTGC-------CCCTGGGAGCCCTGCCCAC 1171

Qy     38 GlyAlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuProThrAlaLeuLeuLeu 57
             :::||||||       ||||||        |||||
Db   1172 CCCCTGGTCTTGCTGCGCCAACCTGCTCAGTGGAAGCTGCTCAAGGGGCTAGCAAGCTG 1231

Qy     58 ValThrThrArgAlaAlaAlaAlaProGlyArgLeuGlyArgLeuGlyArgLeuGlyArgLeuGlnCys 77
             :::|||        |||                              :::
Db   2242 GGTGGGGCAGCTGGCGGCCAGGGCATCTGCGCC---AACTACCTCAAGCTGACCTACTG 2298
```

```
; EARLIER FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 44
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
;   LENGTH: 3382
;   TYPE: DNA
;   ORGANISM: Homo sapiens
US-09-245-281-7

Alignment Scores:
Pred. No.:                                              Length:       3382
Score:                  8.8e-20                         Matches:       172
Percent Similarity:     289.00                          Conservative:   94
Best Local Similarity:  37.84%                          Mismatches:    281
Query Match:            24.47%                           Indels:        158
                        8.98%                            Gaps:           27
DB:                         4
```

US-10-029-347-2 (1-625) x US-09-245-281-7 (1-3382)

```
Qy    1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuAspGlyAlaAspGluLeuProAla  20
Db 1058 CTGCTCGCTTCCCCCACGTGCCCCTTCACCTTGATGGCCTGACGAGCTGCACTCG       1117

Qy   21 ------LeuGlyGlyProGluAlaAlaProCysThrAspProPheGluAlaAlaSer     37
Db 1118 GACTTGGACCTGAGCCGGTGCCTGACAGCTGCTGC-----CCCTGGGAGGCCTGCCCAC  1171

Qy   38 GlyAlaArgVa-LeuGlyGlyLeuLeuSerLysAlaLeuLeuProThrAlaLeuLeuLeu  57
Db 1172 CCCCTGGGGGCTCCAACCTGCTCAGTGGGAAGCGGTGCTCAAGGGGCTAGCAAGCTG    1231

Qy   58 ValThrThrArgAlaAlaAlaProGlyArgLeuArgLysGlnArgLeuCysSerProGnCys  77
Db 1232 CTCACAGCCCGCACA-----------GCATCGAGGTCCGGCCAGTTCCTGCGGAAGAAG  1282

Qy   78 AlaGluValArgGlyPheSerAspLysAspLysProCysTyrPheTyrLysPhePheArg  97
Db 1283 GTGGCTTCTCGGGGCTTCTCCCCCACGCCACCTGCGCGTCCATGCCAGGAGGATGTTCCCC  1342

Qy   98 AspGluArgAlaGluArgAlaTyrArgAlaTyrPheValLysGluAsnGluThrLeuPheAla  117
Db 1343 GAGCGGGCCCTGCTGAGCCACGTGCTGAGGCCCCAACCCCAACCTCTGCAGC      1402

Qy  118 LeuCysPheValProPheValCysTrpIleValCysThrVal------LeuArgGln-  134
Db 1403 CTGTGCTTCGTGTCCCCTTCTGCTGGATCATCTTCCGGTGCTTCCAGCACTTCCGGCT  1462

Qy  135 GlnLeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThrThrSerValTyrLeu  154
Db 1463 GCCTTTGAAGGCTCACCACAGTGCCCGACTGCACGATGACCCTGACAGATGTCTTCTC  1522

Qy  155 LeuPheIleThrSerValLeuSer------  162
Db 1523 CTG---GTCACTGAGGTCCA-TCTGAACACAGGATGCAGCAGCCAGCCTGCAGCGGGAA  1578

Qy  163 ---SerAlaProValAlaAspGlyProArgLeuGlnGly-AspLeuArgAsnLeuCysAr  181
Db 1579 CACACGGGAGCCCAGTGGAGACCTCCAGCGCCGGCCTCCAGGACACTCTGTCTGCTCCGG  1638

Qy  181 gLeuAlaArgGluGlyValLeuGlyValArgArgAlaGlnPheAlaGlnLysGluGlyGlu  201
Db 1639 GGTGGCCCACCCGGCCATGGAAGAGCCCTCTTGTCTTCACCCAGGAGGAGTGCAGGC  1698

Qy  201 nLeuGluLeuArgGlySerLysLysValGlnThrLeuPheLeuSerLysLysValGluPro--  220
Db 1699 CTCCGGGCTCGAGGAGAGGAGACATGGAGCTGGGGCTTCCTG-----CGGGGCTTTGCCGGA  1752

Qy  221 ----G-ValLeuGluThrGluValThrThrTyrGlnPheIleAspGlnSerPheGlnGluPh  239
Db 1753 GCTGGGCCCCGGGGGTGACCAGCAGTCTATGAGTTTTTCCACCTCACCCCTCCAGGCCTT  1812

Qy  239 eLeuAlaAlaLeuSerTyrLeuLeuGluAsp-----------  249
```

```
Qy  383 lArgPheCysArgAspValAlaValLeuSerTyrCysValArgCysProAlaGl    403
Db 2299 CAAGGCCTGCTCGGCCGACTGCGAGGCGGCCTTCCTTCGTCGCATCACTTCCCAAGCG   2358

Qy  403 yGlnAlaLeuuArgLeuIleSerCysArgLeuuValAlaAlaGlnGluLys-LysLysLysSe   423
Db 2359 GCTGCGGCCTAGACCTAGACACACACAATCTC--------                     2389

Qy  423 rLeuGlyLysLysArgLeuGlnAlaLeuSerLeuGlyGlySerSerGlnGlyThrThrLysGl  443
Db 2390 -------------------AACGACTACGGCGTGCGGGAG---                   2410

Qy  443 nLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLe    463
Db 2411 -------CTGCAGCCCTGCTTCAGCCGCCTCAGCT---                       2437

Qy  463 uSerSerLeuThrLeuSerHisCysLysLysLeuProAspAlaValCysArgAspLeuSerG  483
Db 2438 ---GTTCTCGACTCAGCGTAAACCAGATCACTGACGGTGGGTAAAGGTGCTAAGCGA     2493

Qy  483 uAlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGl   503
Db 2494 AGAGCTGACCAAATACAAATTGTGACCTA-TGGGTTTATACAACACCGAATCACCGA    2553

Qy  503 uAlaGlyAlaLeuArgMetLeuSerGluGluGlyValLeuAlaTrpProGnCysArgValGlnThrVa  523
Db 2554 TGCGGAGCCAGTACGTCACCCAAAATCCTG-----GATGAATGCAAAGGCCTCACG--  2635

Qy  523 lArgValGlnLeuProAsp--------ProGlnArgGlyLeuGlnTyrLeu-  537
Db 2606 -CATCTTAAACTGGGAAAAAAACAAAATAACAAGTGAAGGAAGTATTCGCCCGGC    2664

Qy  538 -------ValGlyMet--------  543
Db 2665 TGTGAAGAACAGCAAATCAATCCTGAGGTTGGGATGTGGGCAATCAAGTTGGGGATGA  2724

Qy  541 -------LeuArgGlnSerProAlaLeuThrThrLeuAspLeu-  552
Db 2725 AGGAGCAAAAGCCTTCGCAGAGGCTCTGCGGAACCACCCAGCTTGACCACCCTGAGTCT  2764

Qy  552 uSerGlyCysGlnLeuProAlaProMetValThrThrTyrLeuCysValAlaValValLeuGlnHisGl  572
Db 2785 TGCGTCCAACGGCCATCTCCACAGAAGAAGGAAAGAGCCTTGCGAGGGCCCTGCAG---CA  2841

Qy  572 nGlyCysGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGluGlnSerLeuGl  592
Db 2842 GAACAGCTCCTAGAAAAACTGCGGTGGCTCGACCCAAAATGAACTCAACGATGAAGTGGCAGA  2901

Qy  592 nGluLeuGlnAlaValLysArgAlaLysProAspLeu------ValIleThrHisProAl  610
Db 2902 GAGTTTGCCAGAAATGT--GAAAGTCAACCAGACGTTAAAGCATTTATGGCTTATCCAGA  2960

Qy  610 aLeuAsp 612
Db 2961 ATTCAGAT 2967
```

```
RESULT 12
US-09-245-281-7
; Sequence 7, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
; TITLE OF INVENTION: AND USES THEREOF
; FILE REFERENCE: 07334/118001
; CURRENT APPLICATION NUMBER: US/09/245,281
; CURRENT FILING DATE: 1999-02-05
; EARLIER APPLICATION NUMBER: US 09/207,359
; EARLIER FILING DATE: 1998-12-08
; EARLIER APPLICATION NUMBER: US 09/099,041
; EARLIER FILING DATE: 1998-06-17
; EARLIER APPLICATION NUMBER: US 09/019,942
```

```
Db   1813  CTTACAGCCTTCTTCCTCGTCGCTGCGACGAGGTGGGCACTCAGGAGCTGCTCAGGTT       1872
Qy    250  -----------------G:GGlyValProArgThrAlaAla-----              257
Db   1873  CTTCCAGGAGTGGATGCCCCTCGCGGGGCAGCGACCACGTCCTGCTATCCTCCTTCCT     1932
Qy    258  ---------G:yGlyValGlyThrLeuArgGlyAlaGlnProHisSe                272
Db   1933  CCCGTTCCAGTGCCTGCCTGCAGGGCAGTGGTCCGGCGCGGGAAGAGACCTCTTCAAGAACAGGA   1992
Qy    272  rHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAs    232
Db   1993  TCACTTCCAGTTCACCAACCTCTTCCTGTCGCGGGCTGTTGTCCAAAGCCAAACAGAAA-    2050
Qy    292  pIleGluArgHisPheGlyCysMetValSerGluArgValLysGlnGluAlaLeuArgTr    312
Db   2051  -CTCCTCGGGCATCTGTGTCCCGCGGCAGCCCTGAGCGAGAAAGCCCAAGGCCCTG---TG   2106
Qy    312  p-----------------ValGlGlyLyGlyGlnGlyGlyCysProGlyValAlaProGl   326
Db   2107  GGCACACCTGTT:TCCAGCCTGCGGGGCTACCTGAAGAGCCTGCCCCGCGTTCAGGTCGA   2166
Qy    326  uValThrGluGlyAlaLysGlyLeuGluAspThrGluGlnProGluGluGluGluGluG:    346
Db   2167  AAGCTTCAACCAGGTGCAGG+CATG:--                                    2191
Qy    346  yGluGluGlu:ProAsnTyrProLeuGluLeuLeuTyrCysLeuTyrGluThrGlnGlyAsp-  365
Db   2192  -------CCCACGTTC---ATCTGGATGTGCGCTGCCATCTACGAGACAGAGGCCAGA     2241
Qy    366  -------AlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGlnArgValQ   383
Db   2242  GGTGGGGCAGCTGCGGCGCCAGGGGCATCTGCGCC---AACTACCTCAAGCTGACCTACTG   2298
Qy    383  lArgPheCysArgMetAspValAlaValLeuSerTyrCysValArgCysValArgProAlaGl   403
Db   2299  CAACGGCCTGCTCGGCCGACTCAGCGCGCCCTCTTCCTCGTCGCATCACTTCCCCAAGCG   2358
Qy    403  yGlnAlaLeuAlaArgLeu:leSerCysArgGlyVaLeuValAlaAlaAlaGlnGluLysLysysLysSe   423
Db   2359  GCTGGGCCCTAGACCTAGACAACAACAATCTC---                            2389
Qy    423  rLeuGlyLysArgLeuGlnAl-aSerLeuGlyValGlyGlySerSerGlnGlyThrThrLysGl   443
Db   2390  -----------------AACGACTACGGCGTGCGCGGAG--                       2410
Qy    443  nLeuProAlaSerLeuLeuH:sProLeuPheGlnAlaMetThrArgAspProLeuCysHis:e    463
Db   2411  -----------------CTGCAGCCCTGCTTCAGCCGCGCC*CAC*-                 2437
Qy    463  uSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgAspLeuSerSerGl   483
Db   2438  ---GTTCTCAGACTCAGCGTAAACCAGATCACTGACGCGGTGGGGTAAAGGTGCTAAGCGA   2493
Qy    483  uAlaLeuArgAlaAlaAlaProAlaLeuThrGluLeuLeuGlyLeuLeuAspSerArgLeuSerGl   503
Db   2494  AGAGCTGACCAAATACAAAATTGTGACCTATTTGGGGTTTATACAACAACCAGATCACCGA   2553
Qy    503  :AlaGlyLeuArgMetLeuSerGluGluSerGlyGlyLeuLeuLeuAlaTrpProGlnCysArgValGlnThrVa    523
Db   2554  TGTCGGAGCCAGTAGTCACCAAAATCCTG-----GATGAATGCAAAGGCCTCACG--      2605
Qy    523  lArgValGlnLeuProAsp--------ProGlnArgGlyLeuGlnTyrLeu------       537
Db   2606  -CATCTTAAACTGGGAAAAAACAAAATAACAAGTGAAGGAGGAAGTATCTCCGCCTGGC    2664
Qy    538  -------------------ValGlyMet-----------                        540
Db   2665  TGTGAAGAACAGCAAATCAATCTCTGAGGTTGGGATGTGGGCAATCAGAGTTGGGGATGA    2724
Qy    541  -------------LeuArgGlnSerProAla:euThrThrLeuAspIle                552
Db   2725  AGGAGCAAAAGCCTTGGCAGACGGCTTGCGAGAACCACCCCAGCTTGACCACCCTGAGTCT   2784
```

```
Qy    552  uSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysAlaValAlaValLeuGlnHisGl    572
Db   2785  TGGCG:CCAACGGCCATCCACAGAAGCAGGAGGAGGAAAGAGCCTTGCGAGGGCCCTGCAG---CA   284:
Qy    572  nGlyCysGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGlu:nSerLeuG-      592
Db   2842  GAACACGTCTCTAGAAATACTGTGGCTGACCCAAAATGAACTCAACGATGAAGTGGCAGA    2901
Qy    592  nGluLeuGlnAlaValAlaLysArgAlaLysProAspLeu------ValIleThrHisProAl    610
Db   2902  GAGTTTGGCAGAAATGTT-GAAAGTCAACCAGACGTTAAAGCATTATGGCTTATCCAGA    2960
Qy    610  aLeuAsp       612
Db   2961  ATCAGAT       2967
```

RESULT 23
US-09-207-359B-7
; Sequence 7, Application US/09207359B
; Patent No. 6469140
; GENERAL INFORMATION:
; APPLICANT: Bertin, John
; TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED
; TITLE OF INVENTION: PROTEIN FAMILY AND USES THEREOF
; FILE REFERENCE: 07334-112001
; CURRENT APPLICATION NUMBER: US/09/207,359B
; CURRENT FILING DATE: 1998-12-08
; PRIOR APPLICATION NUMBER: US 09/099,041
; PRIOR FILING DATE: 1998-06-17
; PRIOR APPLICATION NUMBER: US 09/019,942
; PRIOR FILING DATE: 1998-02-06
; NUMBER OF SEQ ID NOS: 47
; SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 7
; LENGTH: 3382
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (245)...(3103)
US-09-207-359B-7

Alignment Scores:
```
Pred. No.:                                          8.8e-20         Length:            3382
Score:                                              289.00          Matches:            172
Percent Similarity:      37.84%                     Conservative:        94
Best Local Similarity:   24.47%                     Mismatches:         281
Query Match:             8.98%                      Indels:             158
Db:                                                 4               Gaps:                27
```

US-10-029-347-2 (1-625) x US-09-207-359B-7 (1-3382)

```
Qy      1  MetLeuAlaGlnProGlnArg:euLeuPheIle:LeuAspGlyAlaAspGluLeuProAla    20
Db   1058  CTGCTGCCGCTTCCCCCACGTGCCCCTTCACCTTCGATGGCTCCGATGGCCTGGAGCCGACGAGCTGACTCG  1117
Qy     21  --------LeuGlyGlyProGluAlaAlaProCysThrArgProPheGluAlaAlaSer     37
Db   1118  GACTTGGACCTGGACCGCGTGCCTGACAGCTCCTGC-----CCCTGGGAGGCTGCCCAC   1171
Qy     38  GlyAlaArgValLeuGlyGlyLeuLeuSerLysAlaLeuLeuGluProThrArgAlaLeuLeu    57
Db   1172  CCCCTGGTCTTCTGGCCAACCTGCTCAGTGGGAGCTGCTCAAGGGGCTAGCAAGCTG    1231
Qy     58  ValThrThrArgAlaAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCys   77
Db   1232  CTCACAGCCCGACA----------GGCATCGAGGTCCCGACCAGTTCCTGCGGAGAAAG   1282
Qy     78  AlaGluValArgGlyPheSerAspLysLysLysTyrPheTyrLysPhePheArg    97
Db   1283  GTGCTTCTCCGGGGGCTTCTCCCCAGCCACCTGGCGCCTATGCCAGGAGGATGTTCCCC    1342
```

```
Qy    98  AspGluArgArgAlaGluAlaTyrArgPheValLysGluGlnThrLeuHisPheAla 117
          :::  |||   ||
Db  1343  GAGCGGGCCCTGCCAGGACCGCCTGCTGAGCCAGCTGGAGGCCAACCCCAACCTGCTGCAGC 1402

Qy   118  LeuCysPheValProPheValCysTrpIleValCysThrVa------------LeuArgGln 134
          ||| |||||    ||||||||:: :::
Db  1403  CTGTGCTTCGTGCCCCCTTCTGCTGGATCATCTTCCGTGTCTTCCAGCACTTCCGTGCT 1462

Qy  135  GlnLeuGluLeuGlyArgAspLeuSerArgThrSerLysThrThrSerValTyrLeu 154
          |||||
Db  1463  GCCTTGAAGGTCACCACAGCTGCCCGACTGCACGATGACCCTGACAGATGTCTTCCTC 1522

Qy  155  LeuPheIleThrSerValLeuSer--------------------- 162
          |||   |||
Db  1523  CTG---GTCACTGGAGGTCCA-TCTGAACAGGATGCAGCCGCCAGCCGGTGCAGCGGGGA 1578

Qy  163  ---SerAlaProValAlaAspGlyProArgLeuGlnGly-AspLeuArgAsnLeuCysAr 181
          :::       |||
Db  1579  CACACGGAGCCCCAGTGGAGACCCTCCACGCCGGCCGGCGACATCTGTGCTGCTGGGCA 1636

Qy  181  gLeuAlaAlaArgGluGlyValLeuValLeuGlyAlaGluLysGlnLeuGluGl------ 201
          ::::|||   |||   |||
Db  1639  GGTGGGCCCACCGGGGCATGGAGAGAGAGAGCCTTTGTCTTCACCCAGGAGGTGCAGGC 1696

Qy  201  nLeuGluLeuArgGlySerLysValGlnThrLeuPheLeuSer-LysGluLeuPro-- 220
          |||      :::||      :::
Db  1699  CTCCGGGCTGCAGGAGAGACATGCAGCTGGGCT-CCTG----CGGGCTTTGGCCGGA 1752

Qy  221  ---GlyValLeuGluThrGluValThrTyrGlnSerPheGlnGluPh------------ 239
          |||   :::
Db  1753  GCTGGGCCCCGGGGGTGACCAGAGTCCTATGAGTTTTCCACCTCACCCTCCAGGCCTT 1812

Qy  239  eLeuAlaAlaLeuSerTyrLeuGluAsp-------------------------- 249
          ::::
Db  1813  CTTTACAGCCTTCTTCCGTCGTTGCTGACGACAGGGTGGCACTCAGGAGCTGCTCAGGTT 1872

Qy  250  ---------GlyGlyValProAlaArgThrAlaAla------------------ 257
          :::
Db  1873  CTTCCAGGAGTGGATGGATGCCCCGTGGGGGCAGCGGCACCGTCTGCTATCCCCTCCT 1932

Qy  258  --------GlyGlyValGlyThrLeuLeuGlyGlyGlyGluAlaGlyAspArgHisSe 272
          ||| 
Db  1933  CCCGTTCCAGTGCTGCAGCGGCCAGTGGTCCGGCGCGGAAGACCTCTTCAAGAACAAGGA 1992

Qy  272  rHisLeuValLeuThrThrArgPheLeuPheGlyLeuLeuSerAlaGluArgMetArgAs 292
          :::
Db  1993  CCACTTCACGTTCACCAACCTCTTCCTCTGTGCCGGCTGCGTGTCCAAGCCAAACAGAAA- 2050

Qy  292  pIleGluArgHisPheGlyCysMetValSerGluValLysGlnGluAlaLeuArgTr 312
          :::|||          ||| 
Db  2051  -CTCCTGCGGCATCTGGTGCCCCGGGCAGACCCGGCCAAGGCCAGGCCCTG---TG 2106

Qy  312  p------------ValGlnGlyGlnGlyGlyGlyCysProGlyValAlaProGl 326
          |||
Db  2107  GGCACACCGTGTTCCAGCCGGCTGGGGGCTACCTGGACGATAGCCTGAACGAGAGCCCGA 2166

Qy  326  uValThrGlyValAlaLysGlyValGluAspThrGluGluGluGluGluGl 346
          :::
Db  2167  AAGCTTCAACCAGGTGCAGCCCATG--------------------------- 2191

Qy  346  yGluGluProAsnTyrProLeuGluLeuTyrCysLeuTyrGluThrGlnGluAsp- 365
          ||| :::
Db  2192  --------CCCACGTTC---ATCTGGATGCTGGCTGCCATCTGCGCC---AAC 2241

Qy  366  ------AlaPheValArgGlnAlaLeuCysArgPheProGluLeuAlaLeuGlrArgVa 383
          :::||| 
Db  2242  GGTGGGCCAGCTGGCGGGCCAGGGGCCATCTGGCCT---AACACCTCAAGCTGACCCACTG 2298

Qy  383  lArgPheCysArgMetAspValAlaValAlaValLeuSerTyrCysSerValArgCysPro 403
          :::|||  |||
Db  2299  CAACGCTGCTCGCCGACTGGCCGGCGCCGCTCCCTCTTCGCGCGCCATCACTCCCCAAGCG 2358

Qy  403  yGlnAlaLeuArgLeuIleSerArgLeuValAlaAlaAlaGlnGluLysGlyLysLysSe 423
          :::||| 
Db  2359  GCTGGCGCCCTAGACCTAGACAACAACAATCTC--- 2389

Qy  423  rLeuGlyLysArgLeuGlnAlaSerLeuLeuGlyGlyGlySerGlnGlyThrThrLysGl 443
          :::
Db  2390  -------------------------AACGACTACGGCGTGCGGGAG- 2410

Qy  443  nLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLe 463
          |||   ||| 
Db  2411  ----------CTGCAGCCCTGCTTCCAGCCGCCCTCACT 2437

Qy  463  uSerSerLeuThrSerHisCysLysLysLysProAspAlaCysArgAspLeuSerGl 483
          ::: |||
Db  2438  ---GTTCTCAGACTCAGGTAAACCAGATCACTGACGGTGGGTAAAGGTGCTAAGCGA 2493

Qy  483  uAlaLeuArgAlaAlaProAlaLeuThrGluLeuGlyLeuLeuHisAsnArgLeuSerGl 503
          ::: |||
Db  2494  AGAGCTGACCAAATACAAAATTGTGACCTATTTGGGTTATACAACAACCAGATCACCGA 2553

Qy  503  uAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrVa 523
          :::
Db  2554  TGTCGGAGCCAGGTACGTCACCAAAATCCTG------GATGAATGCAAGGCCTCACG- 2605

Qy  523  lArgValGlnLeuProAsp---------------ProGlnArgGlyLeuGlnTyrLeu-- 537
          :::
Db  2606  -CATCTTAAACTGGGAAAAACAAAATAACAAGTGAAGGAGGAAGTATCTCGCCCTGGC 2664

Qy  538  -------------------ValGlyMet----------------- 540
          |||||||
Db  2665  TGTGAAGAACAGCAAATCAATCTCTGAGGTTGGGATGTGGGGCAATCAAGTGGGGATGA 2724

Qy  541  ----------------LeuArgGlnSerProAlaLeuThrThrLeuAspLe 552
          |||||||
Db  2725  AGGAGCAAAAGCCTTCGCAGAGGCTCTGCGGAACCACCTGACCACCCTGAGTCTGAGTCT 2784

Qy  552  uSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysValValLeuGlnHisGl 572
          ||| 
Db  2785  TGCGGTCCAACGGCATCTCCAACAGAGGAGAAGAGAGCCTTGCGAGGGCCCTGCAG---CA 2841

Qy  572  rGlyLeuGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGluLeuGl 592
          |||   |||
Db  2842  GAACACGTCCTCTAGAAATACTGTGGCTGCACAAATGAACTCAACGATGAAGTGGCAGA 2901

Qy  592  rGluLeuGlnAlaValLysArgAlaLysProAspLeu------ValIleThrHisProAl 610
          :::
Db  2902  GAGTTTGGCAGAAATGTT-GAAAGTCAACCAGACGTTAAAGCATTTATGGCTTATCCAGA 2960

Qy  610  aLeuAsp 612
          |||
Db  2961  ATCAGAT 2967
```

```
; TYPE: DNA
; ORGANISM: Homo sapiens
; FEATURE:
; NAME/KEY: CDS
; LOCATION: (245)...(3104)
US-09-340-620A-7

Alignment Scores:
Pred. No.:              8.8e-20
Score:                  289.00
Percent Similarity:     37.84%
Best Local Similarity:  24.47%
Query Match:            8.99%
DB:

Length:        3382
Matches:        172
Conservative:    94
Mismatches:     281
Indels:         158
Gaps:            27
```

US-10-029-347-2 (1-625) x US-09-340-620A-7 (1-3382)

```
Db   2785 TGCGTCCAACGGCATCTCCAGAAGAGGAAGAGCCTTGCGAGGGCCCGCAG---CA 2841
          :::     ||| |||||||||
Qy   2785 GTGCTTCCTCCGGGGCTTCTCCCCAGCCACCTGGCGCCGCCATGCCAGGAGGATGTTCCCC 1399

Qy    572 nGlyCysGlyLeuGlnThrLeuSerLeuAlaSerValGluLeuSerGlnSerLeuGl 592
          ||| |    ||| ||
Db     98 AspGluArgArgAlaGluArgAlaTyrArgPheValLysGluAsnGluThrLeuPheAla 117

Db   2842 GAACACGTCTCTAGAAATACTGGCTGACCAAATGAACTCAACGATGAAGTGGCAGA 2901
          :::||||||||||:::||||||||||||
Qy    593 GAGCGGGGCCCTGCAGGACCGCCTGTGAGCCAGCTGGAGGCCAACCCAACCTCTGCAGC 1459

Qy    592 nGluLeuGlnAlaValLysArgAlaLysProAspLeu------ValIleThrHisProAl 610
          ||| |||||||
Db    118 LeuCysPheValProPheValCysTrpIleValCysThrVal------LeuArgGln 134

Db   2902 GAGTTTGGCAGAAATGT-GAAAGTCAACCAGACGTTAAAGCATTTATGGCTTATCCAGA 2960
          :::||||||||:::||||||||
Qy   1460 CTGTGCTCGTGCCCCTTCTGTCTGGATCATCTTCCGGGTGGTTCCAGCACTTCCGTGCT 1519

Qy    610 aLeuAsp 612
          |||||
Db    135 GlnLeuGluLeuGlyArgAspLeuSerArgArgThrSerLysThrThrSerValTyrLeu 154

Db   2961 ATCAGAT 2967
Qy   1520 GCCTTGAGGCTACCACCAGCTGCCCGACTGCCACCGATGACCCTGACAGATGTCTTCCTC 1579


RESULT 15
US-09-245-281-38
; Sequence 38, Application US/09245281
; Patent No. 6369196
; GENERAL INFORMATION:
;  APPLICANT: Bertin, John
;  TITLE OF INVENTION: NOVEL MOLECULES OF THE CARD-RELATED PROTEIN FAMILY
;  TITLE OF INVENTION: AND USES THEREOF
;  FILE REFERENCE: 07334/118001
;  CURRENT APPLICATION NUMBER: US/09/245,281
;  CURRENT FILING DATE: 1999-02-05
;  EARLIER APPLICATION NUMBER: US 09/207,359
;  EARLIER FILING DATE: 1998-12-08
;  EARLIER APPLICATION NUMBER: US 09/099,041
;  EARLIER FILING DATE: 1998-06-17
;  EARLIER APPLICATION NUMBER: US 09/019,942
;  EARLIER FILING DATE: 1998-02-06
;  NUMBER OF SEQ ID NOS: 44
;  SOFTWARE: FastSEQ for Windows Version 4.0
; SEQ ID NO 38
;  LENGTH: 4302
;  TYPE: DNA
;  ORGANISM: Homo sapiens
;  FEATURE:
;  NAME/KEY: CDS
;  LOCATION: (438)...(1184)
;  NAME/KEY: misc_feature
;  LOCATION: (1)...(4302)
;  OTHER INFORMATION: n = A,T,C or G
US-09-245-281-38

Alignment Scores:
Pred No.:                        4302
Score:              3.46e-19     169
Percent Similarity: 285.00       92
Best Local Similarity: 37.66%    278
                       24.39%     156
Query Match:            8.86%     26
DB:                         4

US-10-029-347-2 (1-625) x US-09-245-281-38 (1-4302)

Qy      1 MetLeuAlaGlnProGlnArgLeuLeuPheIleLeuLeuAspGlyAlaAspGluLeuProAla 20
          ::: ||||||||   ||||||   ||||||||||
Db   1115 CTGCTGCGCTTCCCCCACGTGGCCCTCTTCACCTTCGATGCCTGACGAGCTGCCACTCG 1174

Qy     21 ------LeuGlyGlyProGluAlaAlaProCysThrAspCysProPheGluAlaAlaSer 37
          ||| :::    |||
Db   1175 GACTTGGACCTGGACCTGGCCGTGCCTGCCGACAGCTCCTGC-----CCCTGGGAGCCTGCCCAC 1228

Qy     38 GlyAlaArgValLeuGlyLeuSerLysLysAlaLeuLeuProThrAlaLeuLeuLeu 57
          |||  ||  :::||  ||||||||
Db   1229 CCCCTGGCTCCTTGCTGCCAACCTGCTGCTCAGGGGGCTGCTCAGGGGGCTAGCAAGCTG 1288

Qy     58 ValThrThrArgAlaAlaAlaProGlyArgLeuGlnGlyArgLeuCysSerProGlnCys 77
          ::: ||||
Db   1289 CTCACAGCCCCGACA-----GGCATCGAGGTCCGGCCAGTTCCTGCGGAAGAAG 1339

Qy     78 AlaGluValArgGlyPheSerAspLysLysLysLysTyrPheTyrLysPhePheArg 97
          ||| :::    |||||||:::
```

```
Db    2356  CAACGCCTGCTCGGCGGACTGCAGCGCCCCTCCTTCGTCCTGCATCACTTCCCCAAGCG  2415

Qy     403  yGlnAlaLeuArgLeuIleSerCysArgLeuValAlaAlaGlnGluLysLysLysSe     423
                 ||||||   |||  ||    : :             |||
Db    2416  GCTGGCCCTAGACCTAGACAACAACAATCTC----------------------------  2446

Qy     423  rLeuGlyLysArgLeuGlnAlaSerLeuGlyGlyGlySerSerGlnGlyThrThrLysGl  443
                 ||||||                              :::|||
Db    2447  -------------------------------AACGACTACGGCGTCGGGGAG--         2467

Qy     443  nLeuProAlaSerLeuLeuHisProLeuPheGlnAlaMetThrAspProLeuCysHisLe  463
                 ||      |||                      ::||
Db    2468  -------------------CTGCAGCCCTGCTTCAGCCGCCTCACT-------------   2494

Qy     463  uSerSerLeuThrLeuSerHisCysLysLeuProAspAlaValCysArgLeuAspLeuSerGl  483
                     |||                ::: |||||||
Db    2495  ---GTTCTCAGACTCAGCGTAAACCAGATCACTGACGGTGGGTAAAGGTGCTAAGCGA  2550

Qy     483  uAlaLeuArgArgAlaAlaProAlaLeuThrGluLeuGlyLeuGlyLeuGlyLeuSerGl  503
                 :::||||   ||||||||| ||:|||||
Db    2551  AGAGCTGACCAAATACAAAATTGTGACCTATTTGGGTTTATACAACAACCAGATCACCGA  2610

Qy     503  uAlaGlyLeuArgMetLeuSerGluGlyLeuAlaTrpProGlnCysArgValGlnThrVa  523
                 ::::|||||    |:::
Db    2611  TGTCGGAGCCAGGTACGTCACAAATCCTG------GATGAATGCAAAGGCCTCACG--   2662

Qy     523  lArgValGlnLeuProLeuAsp---------ProGlnArgGlyLeuGlnTyrLeu------  537
                 :::||||||
Db    2663  -CATCTTAAACTGGGAAAAACAAAATAACAAGTGAAGGAGGAAGTATCTCGCCCTGGC  2721

Qy     538  --------------------ValGlyMet--------------------             540
                                     ||||||||
Db    2722  TGTGAAGAACAGCAAATCAATCTCTGAGGTTGGGATGTGGGGCAATCAAGTTGGGGATGA  2781

Qy     541  -------------------LeuArgGlnSerProAlaLeuThrThrLeuAspLe       552
                                    |||||||||||||||||||   ||
Db    2782  AGGAGCAAAAGCCTTCGCAGAGGCCTGCGCAGGAACCACCCCAGCTTGACCACCCTGAGTCT  2841

Qy     552  uSerGlyCysGlnLeuProAlaProMetValThrTyrLeuCysValAlaValLeuGlnHisGl  572
                 |:::                           |||||   ||
Db    2842  TGCGTCCAACGGCATCTCCACAGAGAGGAGGAAAGAGCCTTGCGAGGGCCCTGCAG---CA  2898

Qy     572  nGlyCysGlyLeuGlnThrLeuSerLeuAlaSerValGluSerGluGlnSerGluLeuGi  592
                 ||||||   |||    ||
Db    2899  GAACACGTCTCTAGAAATACTGTGCTGACCCAAAATGAACTCAACGATGAAGTGGGCAGA  2958

Qy     592  nGluLeuGlnAlaValLysArgAlaLysProAspLeu  604
                  :::
Db    2959  GAGTTTGGCAGAAATGTT-GAAAGTCAACCAGACGTT  2994
```

Search completed: October  3,  2003, 18:09:07
Job time : 142 secs